



MATLAB在数学建模中的应用

主编 卓金武 副主编 魏永生 秦 健 李必文

北京航空航天大学出版社
BEIHANG UNIVERSITY PRESS

MATLAB 在数学建模中的应用

主　编　卓金武
副主编　魏永生　秦　健　李必文

北京航空航天大学出版社

内容简介

本书从数学建模的角度介绍MATLAB的应用。本书的作者都具有实际的数学建模参赛经历和竞赛指导经验。书中内容完全是根据数学建模竞赛的需要而编排的，涵盖了绝大部分数学建模问题的MATLAB求解方法。全书内容分上下两篇。上篇介绍数学建模中常规方法的MATLAB实现，包括MATLAB交互、数据建模、程序绘图、灰色预测、规划模型等方法；还介绍了各种高级方法的MATLAB实现，包括遗传算法、粒子群算法、模拟退火算法、人工神经网络、小波分析、动态仿真、数值模拟等。下篇以真实的数学建模赛题为案例，介绍了如何用MATLAB求解实际的数学建模问题，给出了详细的建模过程和程序。书中的附件部分介绍了作者在建模竞赛中屡获大奖的经验。相信这些经验对准备参加数学建模竞赛的读者会有所帮助。

本书特别适合作为数学建模竞赛的培训教材或参考用书，也可作为大学"数学实验"和"数学建模"以及"数据挖掘"课程的参考用书，还可作为广大科研人员、学者、工程技术人员的参考用书。

图书在版编目(CIP)数据

MATLAB在数学建模中的应用 / 卓金武主编. -- 北京
：北京航空航天大学出版社，2011.4
ISBN 978-7-5124-0317-8

Ⅰ. ①M… Ⅱ. ①卓… Ⅲ. ①数学模型—计算机辅助计算—软件包，MATLAB Ⅳ. ①TP391.75

中国版本图书馆CIP数据核字(2011)第006288号

MATLAB在数学建模中的应用

主　编　卓金武
副主编　魏永生　秦　健　李必文
责任编辑　董　瑞

*

北京航空航天大学出版社出版发行
北京市海淀区学院路37号(邮编100191)　http://www.buaapress.com.cn
发行部电话：(010)82317024　传真：(010)82328026
读者信箱：bhpress@263.net　邮购电话：(010)82316936
有限公司印装　各地书店经销

*

开本：787×1092　1/16　印张：19.25　字数：493千字
2011年4月第1版　2011年4月第1次印刷　印数：4 000册
ISBN 978-7-5124-0317-8　定价：34.80元

作者简介

卓金武：硕士，2003 年与 2004 年获全国大学生数学建模竞赛一等奖，2007 年获全国研究生数学建模竞赛一等奖，2006 年获全国研究生数学建模竞赛二等奖，2004 年与 2005 年获美国大学生数学建模竞赛二等奖；中国矿业大学数学建模协会创始人之一，并担任第一届数学建模协会执行主席，策划并组织了首届苏北高校数学建模联赛；多次指导学生在全国竞赛中获奖。

魏永生：在读博士，中国工业与应用数学学会会员，获国际一等奖一项，国际二等奖两项，国家一等奖两项，二等奖三项，全国夏令营优秀论文奖一项。曾担任中国矿业大学数学建模协会副主席，苏北数学建模联赛组委会秘书长，第二届苏北数学建模联赛组织者之一。目前，担任北京交通大学大学生数学建模竞赛指导老师，同时创立了北京诺亚数学建模科技有限公司(国内首家提供数学建模咨询服务的公司，网址 http://www.noahmodel.com)。

秦健：硕士，中国工业与应用数学学会会员，获国际二等奖一项，全国大学生数学建模竞赛二等奖一项，全国研究生数学建模竞赛一等奖一项，曾担任中国矿业大学数学建模协会主席，第一届和第二届苏北数学建模联赛组织者之一。现为徐州建筑职业技术学院数学建模教练，所指导的学生在国际和全国大学生数学建模竞赛中获得过一等奖、二等奖若干项。另外，还曾参与苏北数学建模联赛的命题工作，并多次担任苏北数学建模联赛评委。

李必文：学士，毕业于安徽工程大学纺织服装学院，2005 年—2007 年先后在数学建模大赛中荣获安徽赛区二等奖、一等奖和全国一等奖，2007 年获全国大学生书画艺术大奖赛三等奖。曾任安徽工程大学数学建模协会(2004 年安徽省十佳社团)副主席。自 2008 年 3 月，先后担任华硕电脑(苏州)集团 PQA 工程师和南京匹尔睿电子信息技术有限公司数据挖掘工程师。

序　言

全国大学生数学建模竞赛开展近20年了。在全国竞赛的推动下，数学建模教学和竞赛活动已经遍及国内大多数高等院校，各种类型的数学建模教材、参考书、辅导材料层出不穷。这些出版物基本上都出自具有相当教学经验的教师之手。

与这些书的作者不同，我应邀作序并向读者推荐的这本书的编者是一群“80后”。他们不仅多次自己参赛并指导学生参赛，而且都曾是学校建模协会、建模联赛等活动的组织者，他们通过参与竞赛活动自身受益，对数学建模产生了很大兴趣，坚持不懈地从事建模的研究、教学和应用、推广工作。他们的这些经历使得这本书有许多与众不同之处。

我相信，这是这群“80后”们编写并正式出版的第一本书，可能会存在若干不足。但是，编者丰富的建模活动经历反映到书中，使得一些数学知识的应用和处理，一些赛题的分析和编程，特别是附件中关于参赛经验的论述，都具有鲜明的特色，值得准备参加竞赛的读者及数学建模爱好者一读。

姜启源

2010年4月

前　言

CUMCM(全国大学生数学建模竞赛)已成为大学生参赛规模最大的科技竞赛盛事。而MATLAB作为当前最优秀的数学软件，已位列全球最流行软件排行榜的前20名。在CUMCM中能否取得好成绩不仅取决于模型的精妙和合理，还取决于模型的求解。MATLAB在模型的求解方面占有不可或缺的位置。目前，CUMCM设立了“MATLAB创新奖”，也说明了MATLAB在数学建模中的重要地位。

MATLAB入门很容易，但要想精通MATLAB的所有功能，几乎是不可能的，所以针对自己的需要选择性地学习一些MATLAB知识是一种事半功倍的方法。目前，关于MATLAB的书籍很多，关于数学建模的书籍也不少，但将MATLAB与数学建模真正结合在一起的书籍却不多。本书将从数学建模的角度介绍MATLAB的应用。

我们从最初的学生时代自己参加CUMCM和MCM(美国大学生数学建模竞赛)到现在指导学生参加CUMCM和MCM，一路走来，有很多体会。希望通过这本书，将我们对数学建模和MATLAB的学习方法与体会总结出来，让读者快速掌握参加数学建模竞赛应该掌握的数学建模方法和必备的MATLAB编程技术，以提高竞赛成绩。在内容的编排上，也是尽量做到让初学者能够入门，让经验丰富者能耳目一新，思路豁然。书中融入了我们许多参加CUMCM和MCM的真实体会以及多年一线指导学生参加竞赛的经验，使本书具有实用、创新和鲜活的特征。

全书分上下两篇讲述MATLAB在数学建模中的应用。上篇中第1～3章讲的是数学建模中常规方法的MATLAB实现，读者看完这部分后，就可以用MATLAB解决数学建模中的一些基础问题了，如数据处理、程序绘图、求解目标规划模型、数据量较少时的预测等。随着数学建模竞赛的发展，题目对参赛者创造性解决问题的要求越来越高。也就是说，参赛者需要有自主开发程序的能力，才有可能在比赛中取得好成绩，所以就要学习更为复杂的编程方法和技巧。于是在上篇剩下的篇幅安排了数学建模高级方法的MATLAB实现，包括遗传算法、粒子群算法、模拟退火算法、神经网络、小波分析、动态模拟、数值模拟等。通过这些内容的学习，读者基本上就可以求解一些复杂的问题了。现在的建模竞赛经常会出现的问题不是模型建不出来，而是模型求解不出来，所以高水平的MATLAB编程和对常用方法的熟悉将对参加数学建模竞赛相当有帮助。下篇介绍了数学建模竞赛中典型真题的MATLAB求解方法及过程，通过对一些真题案例的解析，加强读者对MATLAB的驾驭能力，巩固读者对MATLAB的操作能力，同时增加读者对数学建模问题的敏感度。

纵观全书，可发现该书的特点相当鲜明，主要表现在如下几个方面：

(1) 结构编排合理。在讲解方法时，由浅入深，循序渐进，让初学者知道入门的切入点，让建模和编程高手又有借鉴的空间。书中有很多小程序可供初学者套用及学习，又不乏实现功能复杂的篇幅较长的程序，这部分主要面向学有余力的读者。

(2) 既突显数学建模思想又包含了完整的MATLAB程序。目前，现有的数学建模书籍完整附带程序的较少，而本书给出了大量实用、完整、详细的源代码，这对于初学数学建模和MATLAB的读者来说，具有非常大的参考价值。

(3) 部分新型方法是同类书籍中比较少见的。本书中的动态仿真和小波分析是目前其他建模书籍中比较少见的。动态仿真，美国大学生刚开始尝试在建模比赛中使用，在国内尚属空

白；小波分析虽有些难懂，但确实是一种有力的工具，在本书中也对其进行了深入浅出的介绍。

（4）本书涉及知识面宽广，以建模为中心，辐射计算机、生物学、环境科学、心理学、古典文学、影视、服装设计以及艺术学。比如，将污染物扩散的四维水质模型与服装设计面料的透气透湿性能联系起来，找出信息隐藏的数字水印技术与《孙子兵法》中的“出其所不趋，趋其所不意”的思想结合，等等。书中所涵盖的这些领域虽然只是被略微带过，但从侧面已折射出数学建模是广泛服务于社会各个领域的。在现代社会，某学科单打独斗的时代已经过去了，本书在无形之中已经树立了一个意识：知识之间是相通的。

（5）附件中关于参加建模竞赛的经验体会也在同类书籍中比较少见，也都是我们参加数学建模竞赛的真实体会。我们站在一个成功参赛者的角度与读者进行交流，更容易让读者感同身受。比如，书中介绍了数学建模中的项目管理和时间管理，这些都是容易被参赛队员忽视却对竞赛成绩影响相当大的因素。再比如，书中介绍了目标建模方法，这是一种非常实用的数学建模方法。有了这些经验和实用方法的指导，读者会有的放矢地去准备数学建模竞赛，同时也能增强读者参加数学建模竞赛的信心。

（6）书中内容以数学建模竞赛为主线，同时有意将数学建模向实际的生产力方向推广，真正体现了数学建模的魅力。这具体体现在第 2 章规划问题所涉及的投资问题，第 10 章彩票问题以及第 14 章出版社资源配置问题等内容的编排上。建模如果只是单纯的论文，不能产生社会价值，那么建模永远只会停留在学术层面，不能创造社会价值就等于没有价值。所以纵观通篇书稿，本书都在有意引导读者思考如何让建模产生更实际的社会价值。

（7）书中原创知识占据绝大部分版面，让读者少了几分似曾相识或模型雷同的感受。常规方法、高级方法和真题案例的结构部署也让本书独树一帜，让读者在学习 MATLAB 和数学建模过程中有一个循序渐进的过程，使读者在短时间内成为一个 MATLAB 程序开发高手，同时成为一个数学建模高手。

（8）内容详实，文字活泼，字里行间透射灵动鲜活的魅力。在本书编写过程中，在保证描述精准的前提下，我们摒弃了那些刻板、索然无味的文字，让文字也有了活力。

感谢清华大学姜启源教授在百忙之中指导本书的修改并为本书写序，同时感谢中国矿业大学的张兴永教授、周圣武教授，安徽工程大学田丽教授、王传玉教授对我们的培养和指导。在本书的编写过程中，中国科学院的王恺博士对第 10 章和第 12 章进行了修改，天津科技大学的谢中华老师关于书稿的修改提出了众多建设性意见，此外马国庆、李小波、王欣、刘甫、李洋等也给予了很多帮助，在此向他们表示感谢。

书中所涉及的相关程序源代码及历年数学建模赛题，请读者登录北京航空航天大学出版社网站（http://www.buaapress.com.cn/）免费下载。同时，北京航空航天大学出版社联合 MATLAB 中文论坛（http://www.iLoveMatlab.cn/）为本书设立了在线交流板块，网址 http://www.iLoveMatlab.cn/forum-198-1.html，作者有问必答，并提供系列增值服务！

本书适合作为数学建模竞赛和 MATLAB 应用的培训教材或参考用书，也可作为大学“数学实验”、“数学建模”以及“数据挖掘”课程的参考用书，还可作为广大科研人员、学者、工程技术人员的参考用书。

由于时间仓促，加之作者水平有限，错误和疏漏之处恳请广大读者批评指正。

本书勘误网址为：http://www.iLoveMatlab.cn/thread-114793-1-1.html。

编　者

2010 年 5 月

目　录

上篇

方法演绎

MATLAB 是当前最优秀的数学软件，随着其版本的不断升级，其强大的功能得到了进一步的加强和完善。数学建模中的模型和方法很多，但最常用的就几种，本篇将介绍这几种常用方法的 MATLAB 实现方式（下表给出了这些方法的应用场景和本书中关于这些方法的特色之处）。

方法名称	在数学建模中的应用场合	特色内容
常规方法	数据处理（数据清预处理、数值计算、数据拟合）、绘制图形、建议预测	MATLAB 交互、薄膜渗透率模型的搭建、地形地貌图形的绘制以及 2002 年真题光源投影区域的绘制都会给初学者耳目一新的感觉
规划问题解法	多约束线性规划、整数规划和不太复杂的多约束非线性规划、整数规划	所选用案例浅显易懂，生动有趣
灰色预测	数据量少的情况下预测	案例皆是全国性竞赛真题，设计的程序容易套用
遗传算法	①求解多约束规划模型； ②训练人工神经网络	①对遗传算法计算过程进行了升级，令人耳目一新； ②将弗洛伊德梦的仿同理论与遗传算法结合起来，找出两者之间契合点； ③内容巧妙编排，透彻易懂
粒子群算法	①求解无约束多元非线性规划模型； ②训练人工神经网络	①对真实鸟群觅食进行计算机模拟； ②首次提出体育裁判神经网络模型； ③对 BP 算法能够寻优到极值的原理进行了通俗易懂的叙述，理论简洁而深刻； ④PSO - BP 神经网络使用了 MATLAB 源程序，没有使用神经网络工具箱，让读者“知其然并知其所以然”
模拟退火算法	①经典 TSP、背包问题； ②求解复杂多约束非线性规划模型	跟踪当前国际模拟退火算法最新研究动态，下载 ASA 程序包并深度解读，让初学者能入门，让熟知该算法的读者思路豁然开朗

续表

方法名称	在数学建模中的应用场合	特色内容
人工神经网络	数学建模中一切聚类、评价及预测模式的问题	①在理论上并列解析 BP 网络和 RBF 网络； ②从全新角度阐述神经网络的基本原理，避免与其他同类书籍雷同，偏重原创和细致； ③设计四层径向基小波神经网络结构； ④同时用 MATLAB 神经网络工具箱和开发 MATLAB 源程序对公路运量问题进行预测
小波分析	①海量数据趋势挖掘； ②组建小波神经网络	①多角度切入小波神经网络； ②引入融合数学建模思想和小波分解思想的数字水印技术
计算机虚拟	①模型难以求出解析解； ②动态(动画)展现数学模型； ③复杂非线性规划问题粗略求解	①有趣解说 π 求解的蒙特卡罗模拟； ②动态仿真表现力强，但是当前具有动态仿真的数学建模书籍还不多见； ③用四种蒙氏模拟法解决非线性规划； ④电影《功夫熊猫》武打片段电影仿真； ⑤完整地组建、求解四维水质模型，并可用计算机模拟连续的瞬间水质情形

第 1 章

数据建模常规方法的 MATLAB 实现

数据处理是数学建模的基础，通常遇到的问题是对采集到的数据进行处理和分析，从而得到这些数据所反映的信息。从数学建模的角度，将数据反映出来的信息转化成数学表达式是建模的基础，所以通常对数据的处理就是趋势分析和将数据转化为函数表达式，这里通常用到的是数据的插值和曲线拟合。应该说，MATLAB 的应用使得这些方法很容易实现，尤其当数据量较大时，MATLAB 的数据处理优势就更明显了。

1.1 数据的读入与读出

用户用 MATLAB 进行科学计算时不可避免地要用到大量的数据，而方便的数据处理方法会让用户更得心应手。MATLAB 提供了多种处理数据的方法，一种方法是将数据输出，然后复制粘贴到其他软件中进行处理，但是这种方法不方便；另一种方法是与 Excel 和记事本进行数据交互，这种方法方便了数据的操作。

1.1.1 Excel 与 MATLAB 的数据交互

MATLAB 中的 Excel Link 工具是一个实现与 Excel 进行交互的插件。通过连接 Excel 和 MATLAB，用户可以在 Excel 工作表空间和宏编程工具中使用 MATLAB 的数值计算和图形处理等功能，不需要脱离 Excel 环境。同时由 Excel Link 保证两个工作环境中的数据交换和同步更新。

Excel Link 的安装：首先，要确保系统中正确安装了 Excel 软件、MATLAB 和 Excel Link 工具箱。启动 Excel，选择菜单“工具”项下的“加载宏”项，弹出一个加载宏对话框，选中 Excel Link 项（一般是 Spreadsheet Link EX for use with MATLAB）。如果该项不存在，则通过浏览目录，在目录 MATLAB\toolbox\exlink 文件夹下找到 excllink. xla 文件，单击“确定”，选中 Excel Link 项并确定后，在 Excel 中多了一个 Excel Link 工具条，如图 1－1 所示。

startmatlab putmatrix getmatrix evalstring getfigure wizard preferences

图 1－1 Excel Link 工具条

经过以上的设置后就可以使用 Excel Link 了，读者看到工具条后会很容易学会其操作，这里不再赘述，有问题可参考 Help。

1.1.2 记事本与 MATLAB 的数据交互

在 MATLAB 中可以读取存入记事本中的数据，也可以将内存中的变量数据保存到记事本中，以便后期操作。

1. MATLAB 读取记事本中的数据

当记事本中记录的全部都是数据时，可以用函数 load，其形式是：

load(‘filename. * * *’)

其中，文件扩展名可以是任意记事本文件的扩展名，运行此函数后就会把记事本中的数据按矩阵的形式放入名为 filename 的变量中。

当记事本中的数据结构变得复杂时，例如表 1－1 中的数据(第 1 行除外)，就不能用函数 load 了，此时函数 textread 是最优的选择，其规则是：

[A,B,C,…]=textread(‘filename’,‘format’,N)

其中，A,B,C,…为每一列数据将要保存的变量名；format 为读取格式；N 为读取次数。

例如上面的文件保存为当前路径下的 t. txt 文件后，输入命令得到：

```
>> [name,type,x,y,answer] = textread('t.txt','%s Type%n %f %n %s',2)
name =
    'Sally'
    'Tom'
type =
     1
     1
x =
    3.1000
    2.5000
y =
    45
    20
answer =
    'Yes'
    'No'
```


表 1－1 数 据

names	types	x	y	answer
Sally	Type1	3.1	45	Yes
Tom	Type1	2.5	20	No
Joy	Type3	0.51	1	No
Lily	Type2	2.1	80	Uncertain

在命令中‘%s Type%n %f %n %s’即为 format 的内容，最后一个参数 2 表示读取 2 行数据。

在 MATLAB 中，还有一个函数 fscanf 可以读取记事本中的文件，其功能更加强大，如有需要，请直接在 Help 中寻求帮助。

2. 将 MATLAB 数据写入记事本中

同样，如果只需要保存数据，并且以后也只用 MATLAB 调用，可以用 save 命令，其语法是：

save file obj1 obj2…

将各变量 obj1,obj2…存入文件 file 中(为. mat 格式)，用户无法直接查看文件中保存

的数据。

如果想将数据保存为平常的.txt文档,可以由其他软件进行读取,函数fprintf是一个不错的选择。还以上面的数据为例,要将这些量按格式保存在一个tp.txt文档中,可以建立以下命令文件。

```
fid = fopen('tp.txt','wt');
fprintf(fid,'This is the database of class 1.\n');
name = 'Sally'; types = 1; x = 3.1; y = 45; answer = 'Yes';
fprintf(fid,'%s Type%u %f %u %s \n',name,types,x,y,answer);
name = 'Tom'; types = 1; x = 2.5; y = 20; answer = 'No';
fprintf(fid,'%s Type%u %f %u %s \n',name,types,x,y,answer);
name = 'Joy'; types = 3; x = .51; y = 1; answer = 'No';
fprintf(fid,'%s Type%u %f %u %s \n',name,types,x,y,answer);
name = 'Lily'; types = 2; x = 2.1; y = 80; answer = 'Uncertain';
fprintf(fid,'%s Type%u %f %u %s \n',name,types,x,y,answer);
fclose(fid);
```

运行后,就会建立一个tp.txt文档,其中的内容如下:

```
This is the database of class 1.
Sally Type1 3.100000 45 Yes
Tom Type1 2.500000 20 No
Joy Type3 0.510000 1 No
Lily Type2 2.100000 80 Uncertain
```

现在对程序中的语句进行说明。第1段是打开一个文件,返回一个指标fid,第2段是直接在文件中写入一句话“This is the database of class 1.”,然后就是每赋值一次,就向文件中写入一次:

```
fprintf(fid,'%s Type%u %f %u %s \n',name,types,x,y,answer);
```

其格式为

```
fprintf(id,'format',arg1,arg2,…);
```

其中,%s表示字符串(string);%f表示浮点数(float);%u表示十进制数。

另外还有很多格式,用时请参阅Help中的fprint以及其下的formatting strings。

在这里还要注意,不能用函数textread读取后再用函数fprintf进行写操作。这是因为函数textread读取数据后保存为cell结构,而函数fprintf不能进行cell结构数据的写操作。

1.2 数据拟合方法

曲线拟合也称为曲线逼近,它与插值函数有些区别,只要求拟合的曲线能合理地反映数据的基本趋势,而并不要求曲线一定通过数据点。曲线拟合有几种不同的判别准则,如使偏差的绝对值之和最小、使偏差的最大绝对值最小和使偏差的平方和最小(即最小二乘法)。常用的方法是最后一种。

1.2.1 多项式拟合

将数据点按多项式的形式进行拟合，使用最小二乘法可以确定多项式的系数。多项式拟合有指令语句和图形窗口两种方法。

1. 多项式拟合指令

polyfit(X,Y,N)：多项式拟合，返回降幂排列的多项式系数。

polyval(P,xi)：计算多项式的值。

其中，X,Y 是数据点的值；N 是拟合的最高次幂；P 是返回的多项式系数；xi 是要求的点的横坐标。

例如要对表 1-2 中的数据进行多项式拟合。

表 1-2 实例数据

x	1	2	3	4	5	6	7	8	9
y	9	7	6	3	−1	2	5	7	20

拟合的命令如下：

```
x=[1 2 3 4 5 6 7 8 9];
y=[9 7 6 3 -1 2 5 7 20];
P=polyfit(x,y,3);
xi=0:.2:10;
yi=polyval(P,xi);
plot(xi,yi,x,y,'r*');
```

得到的拟合曲线与原始数据的对照如图 1-2 所示。

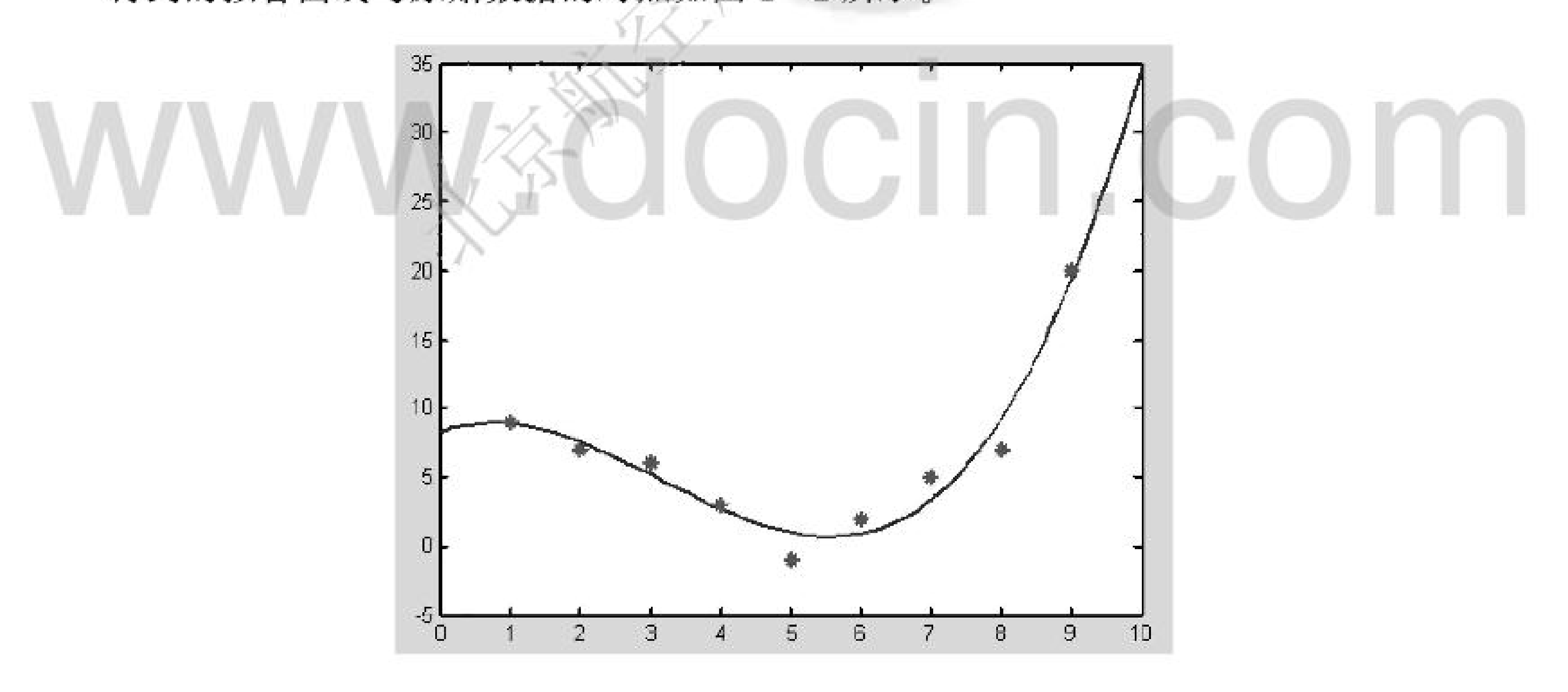


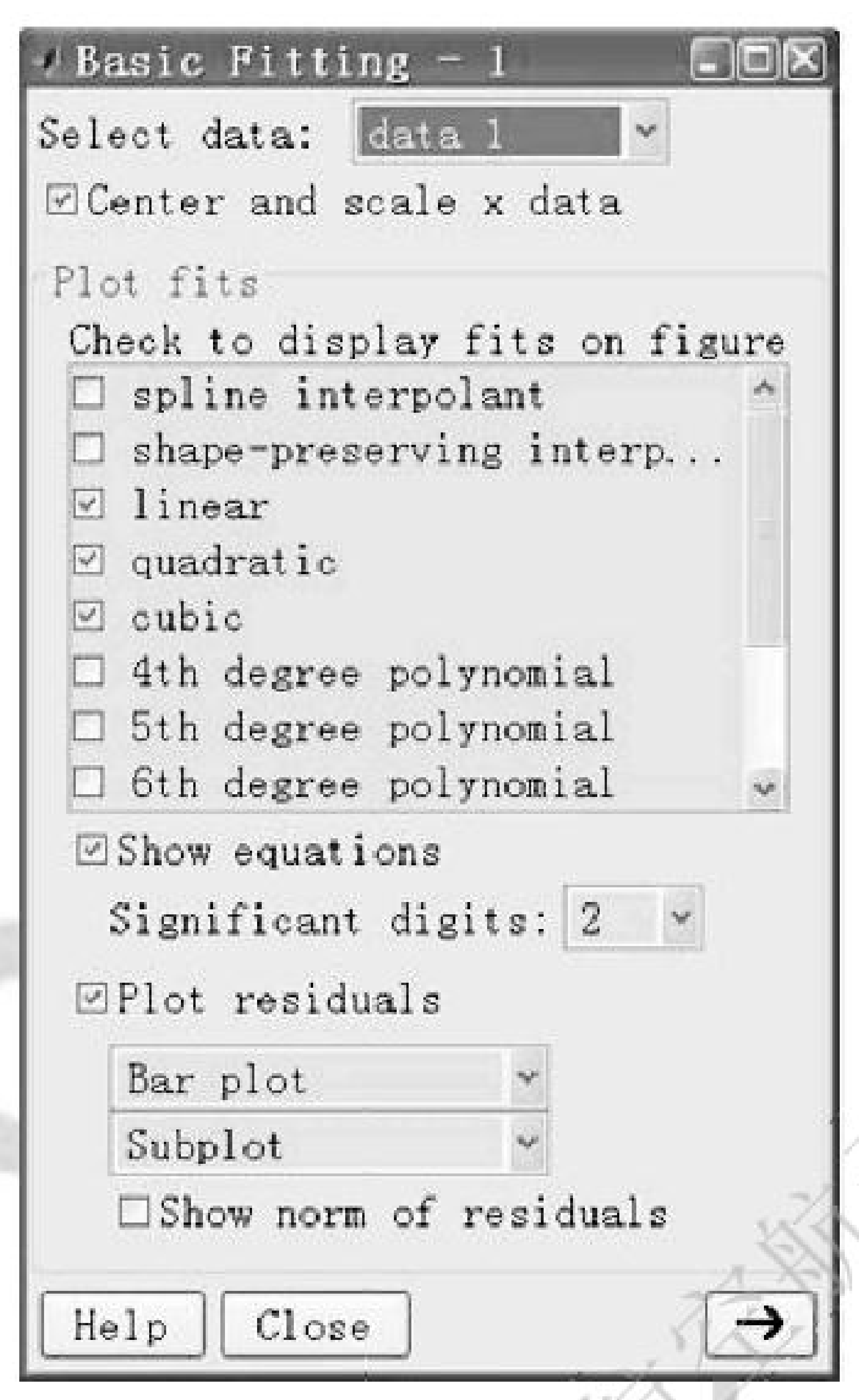
图 1-2 拟合曲线与原始数据的对照

2. 图形窗口的多项式拟合

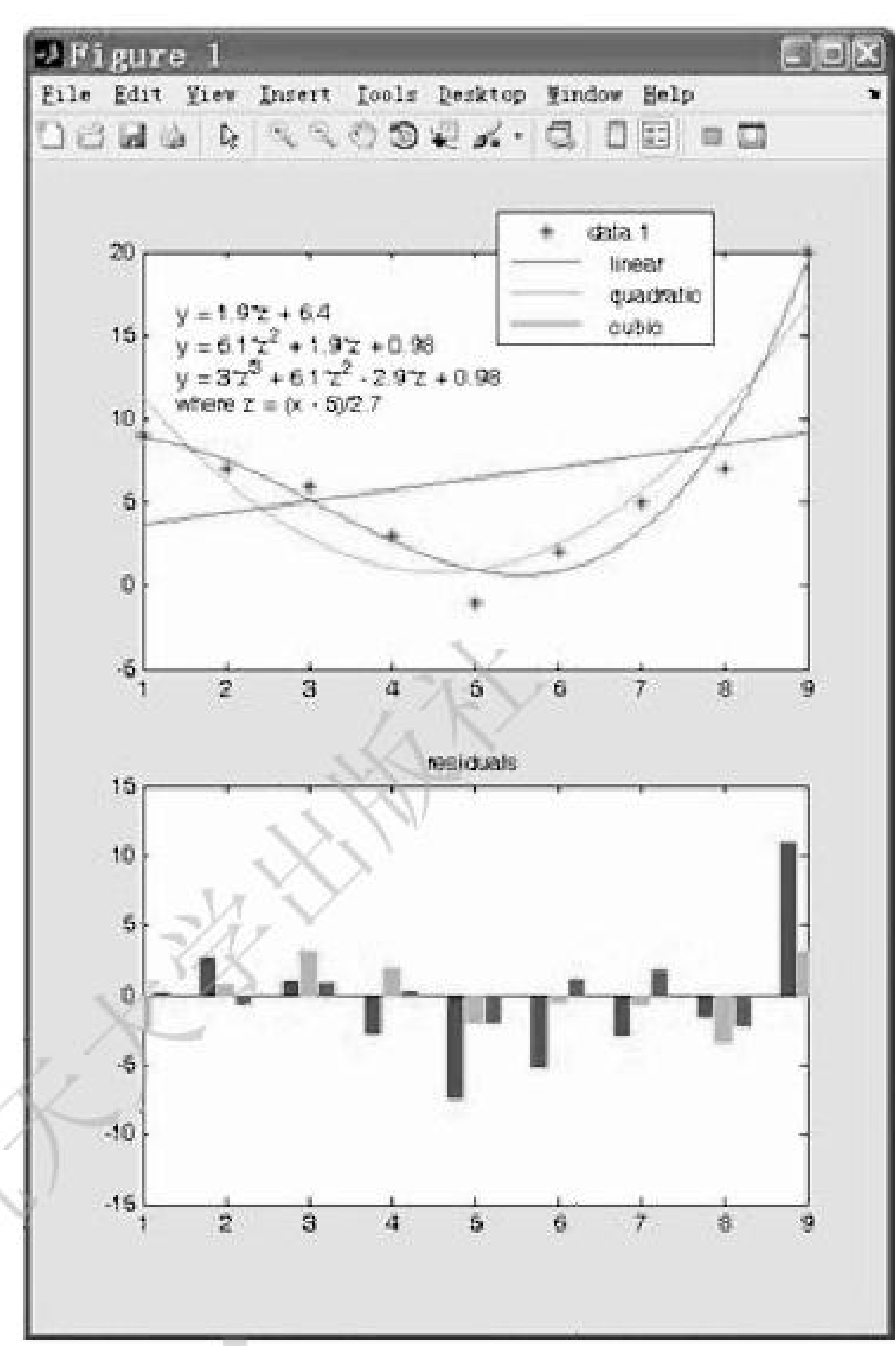
在图形窗口中可以用菜单方式对数据进行简单、快速、高效的拟合。以表 1-2 中的数据为例，先画出数据点：

```
>> x=[1 2 3 4 5 6 7 8 9];y=[9 7 6 3 -1 2 5 7 20];plot(x,y,'r*');
```

然后在图形窗口单击 Tools→Basic Fitting，打开对话框，并按图 1-3(a)进行操作，就可以得到图 1-3(b)的效果，其中分别用线性、二阶、三阶对数据进行了多项式拟合，下面的柱状图显示残差，可以看出，进行三阶多项式拟合的效果是最好的，与前述结果一样。



(a)



(b)

图 1-3　图形窗口的多项式拟合

1.2.2　指定函数拟合

在 MATLAB 中也可以用用户自定义的函数进行拟合。通过下面的例子可以了解用指定函数进行数据拟合的基本方法。

在某次阻尼振荡实验中测得如表 1-3 所列的 18 组数据点。

表 1-3　数据点的坐标

x	0	0.4	1.2	2	2.8	3.6	4.4	5.2	6
y	1	0.85	0.29	−0.27	−0.53	−0.4	−0.12	0.17	0.28
x	7.2	8	9.2	10.4	11.6	12.4	13.6	14.4	15
y	0.15	−0.03	−0.15	−0.071	0.059	0.08	0.032	−0.015	−0.02

在 MATLAB 中可以用如下命令画出上述点的散点图(见图 1-4)：

```
x = [0;0.4;1.2;2;2.8;3.6;4.4;5.2;6;7.2;8;9.2;10.4;11.6;12.4;13.6;14.4;15];
y = [1;0.85;0.29; - 0.27; - 0.53; - 0.4; - 0.12;0.17;0.28;0.15; - 0.03; - 0.15; - 0.071;0.059;
0.08;0.032; - 0.015; - 0.02];
plot(x,y,'r*');
```

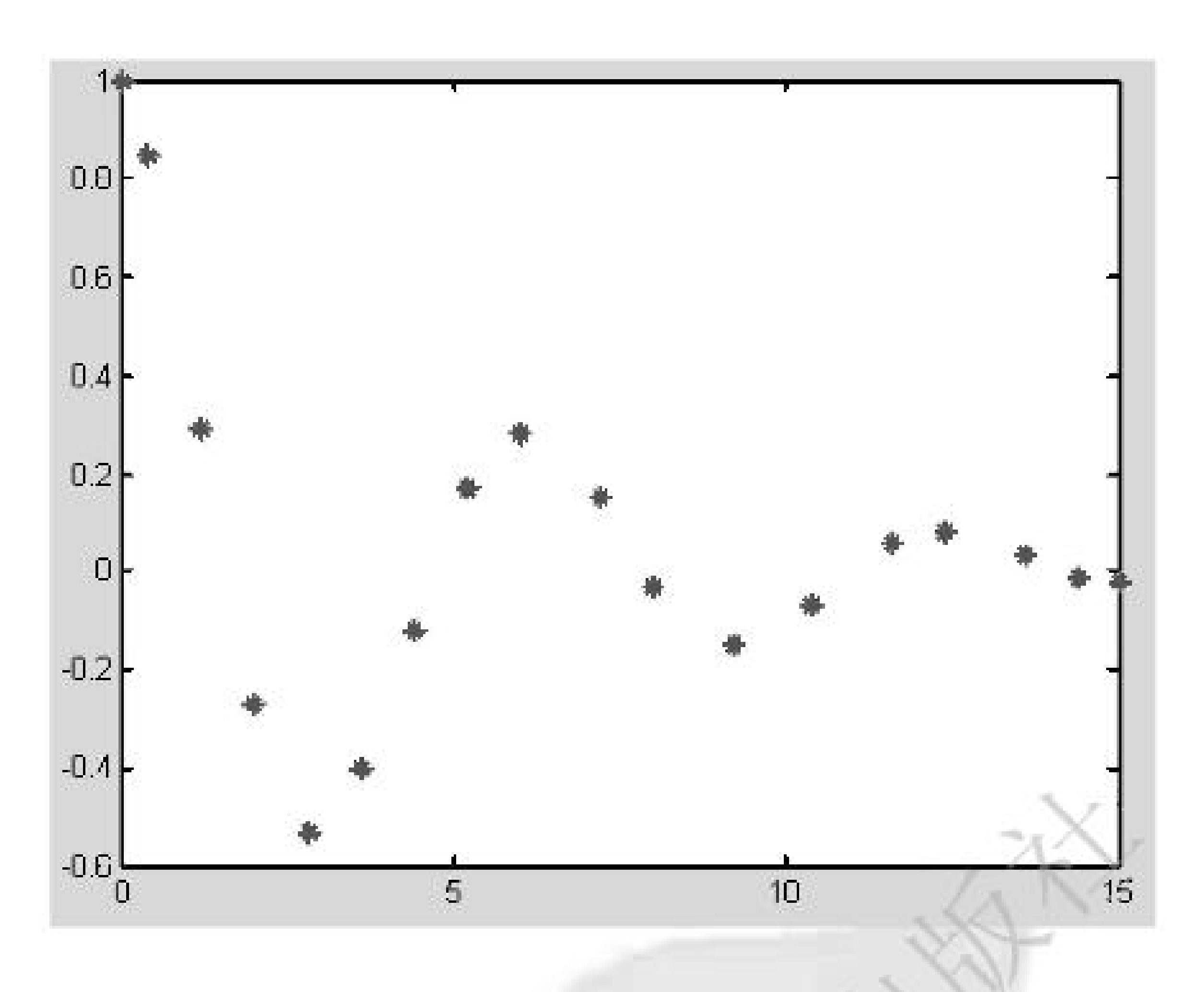


图 1-4　数据散点图

知道其对应的函数形式为 $f(t)=a\cos(kt)\mathrm{e}^{wt}$，则可用 MATLAB 进行拟合。编写如下的 M 文件：

```
syms t
x = [0;0.4;1.2;2;2.8;3.6;4.4;5.2;6;7.2;8;9.2;10.4;11.6;12.4;13.6;14.4;15];
y = [1;0.85;0.29; - 0.27; - 0.53; - 0.4; - 0.12;0.17;0.28;0.15; - 0.03; - 0.15; - 0.071;0.059;
0.08;0.032; - 0.015; - 0.02];
f = fittype('a * cos(k * t) * exp(w * t)','independent','t','coefficients',{'a','k','w'});
cfun = fit(x,y,f)        % 显示拟合函数
xi = 0:.1:20;
yi = cfun(xi);
plot(x,y,'r*',xi,yi,'b-');
```

运行此程序，在 Command Window 可得到如下的运行结果(图形结果如图 1-5 所示)：

```
Warning: Start point not provided, choosing random start point.
 > In fit > handlewarn at 714
   In fit at 328
cfun =
     General model:
       cfun(t) = a * cos(k * t) * exp(w * t)
     Coefficients (with 95 % confidence bounds):
       a =       0.9987   (0.9836, 1.014)
       k =        1.001   (0.9958, 1.006)
       w =      - 0.2066   ( - 0.2131, - 0.2002)
```

程序中，函数 fittype 是自定义拟合函数；cfun=fit(x,y,f)是根据自定义的拟合函数 f 来拟合数据 x，y。注意：此处数据必须为列向量的形式。函数 fittype 和函数 fit 的用法与参数

规则可参考 Help。

从结果可以看出，拟合的曲线为 $f(t)=0.9987\cos(1.001t)e^{-0.2066t}$。拟合曲线给出了数据的大致趋势，结果很好，并给出了各参数的置信区间。

注意：Command Window 中给出的 Warning，是由 a，k，w 3 个参数的初始值未给出导致的，因此如果拟合结果不理想，可以多运行几次。

1.2.3 曲线拟合工具箱

MATLAB 提供了曲线拟合工具箱，其功能很强大，使用也很方便。依次单击 Start→Toolboxes→Curve Fitting→Curve Fitting Tool（cftool）就可以打开曲线拟合工具箱，或者在 Command Window 中直接输入 cftool 命令打开。打开后的界面如图 1－6 所示。

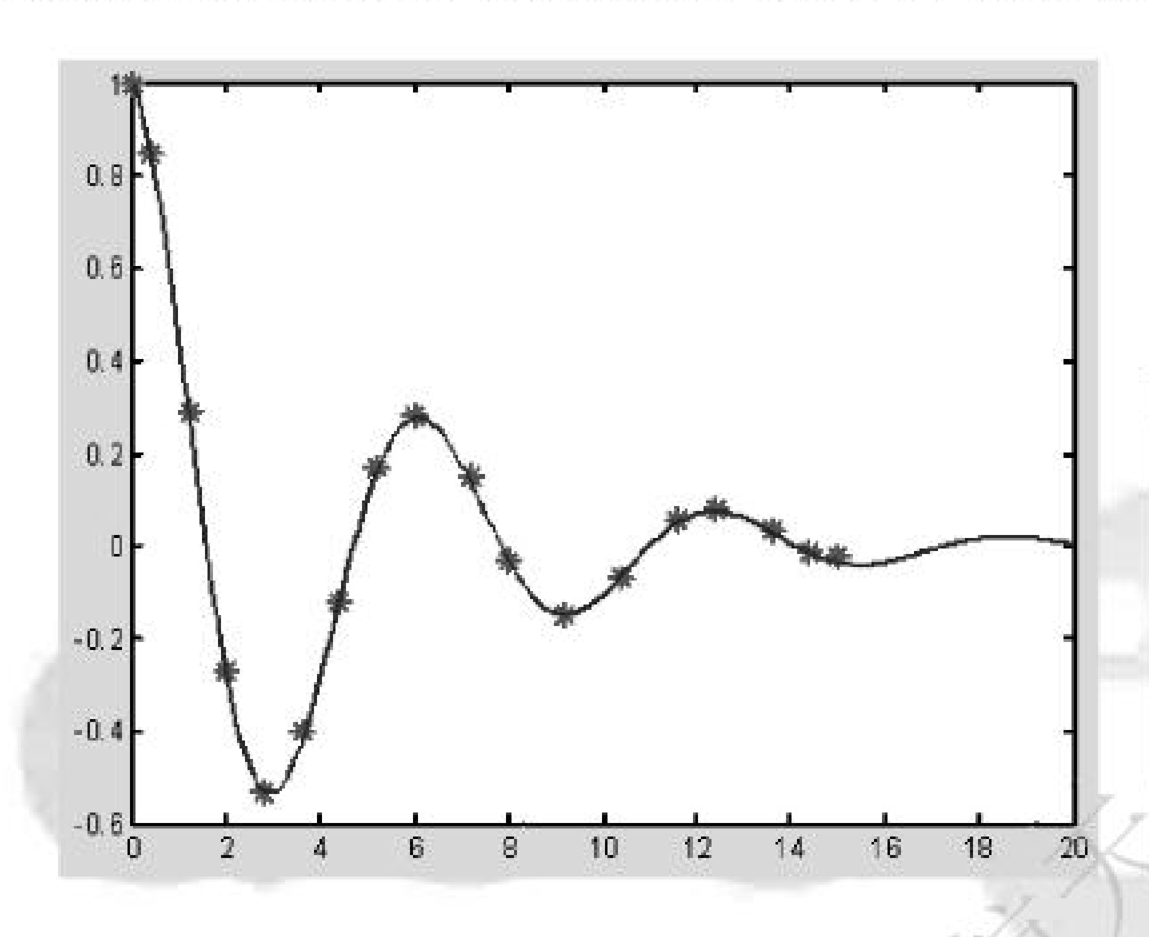


图 1－5 程序运行结果

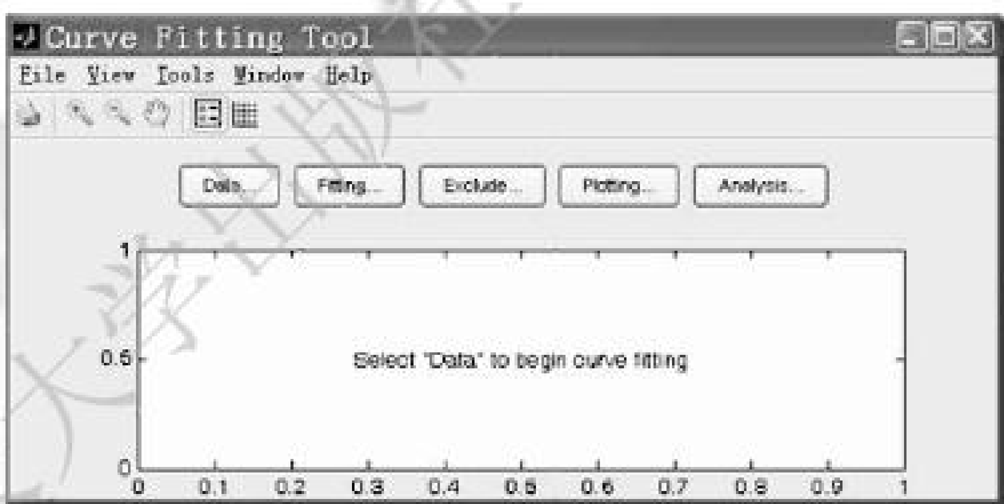


图 1－6 曲线拟合工具箱界面

界面中 5 个按钮的功能分别如下：

Data：输出、查看和平滑数据。

Fitting：拟合数据、比较拟合曲线和数据集。

Exclude：可从拟合曲线中排除特殊的数据点。

Plotting：选定区间后，单击按钮可选择原始数据和拟合的数据作图。

Analysis：对拟合进行满意度、偏差等分析。

曲线拟合工具箱中包含了各种常用的数据拟合方法，可以对各种函数进行拟合，具体的操作方法请参考 Help。

1.3 数据拟合应用实例

1.3.1 人口预测模型

表 1－4 是某地区 1971—2000 年的人口数据，试给出该地区人口增长的数学模型。根据表中的数据做出如图 1－7 所示的散点图。

从图 1－7 可以看出，人口随时间的变化呈非线性过程，而且存在一个与横坐标轴平行的渐近线，故可以用 Logistic 曲线模型进行拟合。

表 1-4　某地区人口变化数据

年　份	时间变量 t=(t 年份－1970)	人口 y/人	年　份	时间变量 t=(t 年份－1970)	人口 y/人
1971	1	33 815	1986	16	34 520
1972	2	33 981	1987	17	34 507
1973	3	34 004	1988	18	34 509
1974	4	34 165	1989	19	34 521
1975	5	34 212	1990	20	34 513
1976	6	34 327	1991	21	34 515
1977	7	34 344	1992	22	34 517
1978	8	34 458	1993	23	34 519
1979	9	34 498	1994	24	34 519
1980	10	34 476	1995	25	34 521
1981	11	34 483	1996	26	34 521
1982	12	34 488	1997	27	34 523
1983	13	34 513	1998	28	34 525
1984	14	34 497	1999	29	34 525
1985	15	34 511	2000	30	34 527

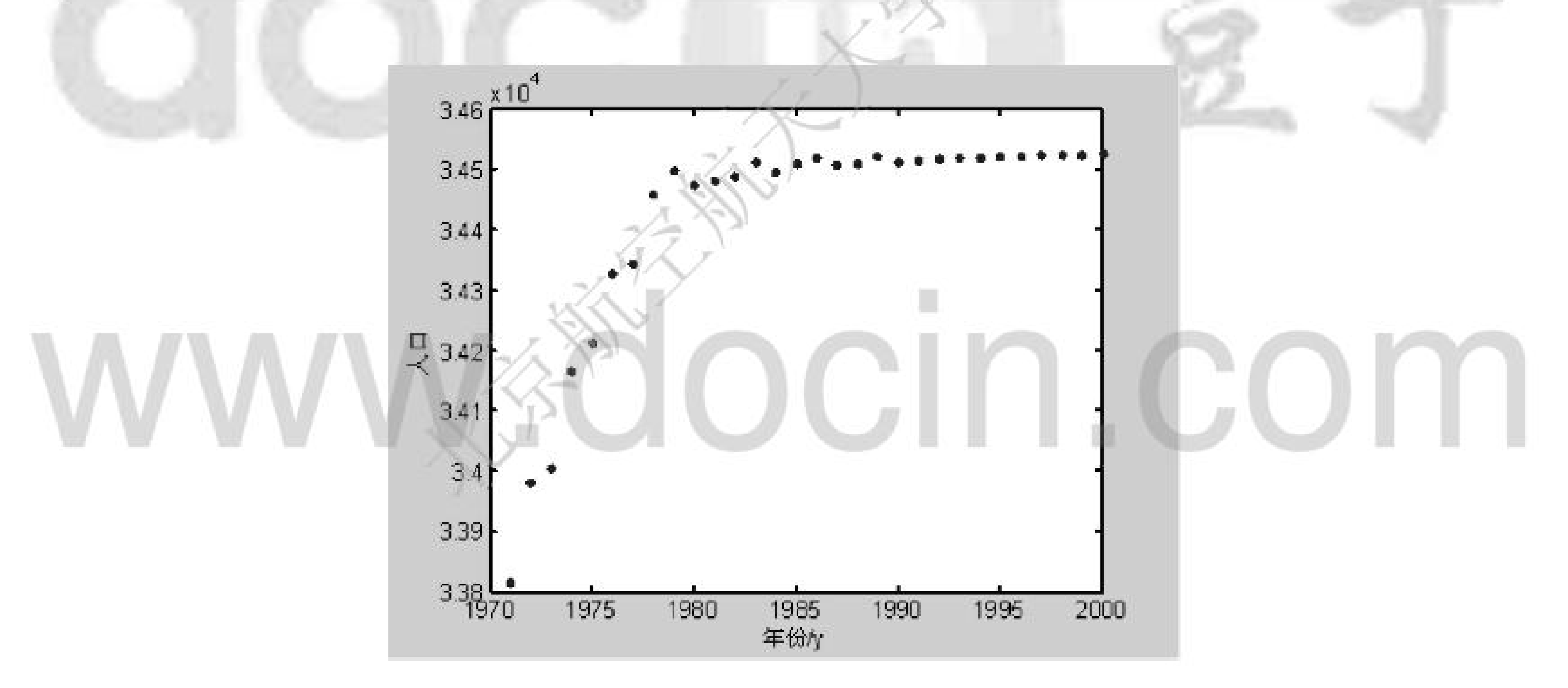


图 1-7　某地区人口随时间变化的散点图

因为 Logistic 曲线模型的基本形式为

$$y=\frac{1}{a+b\mathrm{e}^{-t}}$$

所以，只要令 $y'=\dfrac{1}{y}$，$x'=\mathrm{e}^{-t}$，就可以将其转化为直线模型：

$$y'=a+bx'$$

下面分别用 MATLAB 进行回归分析拟合计算。首先，编写 MATLAB 回归拟合程序，见P1-1。

程序编号	P1-1	文件名称	Population.m	说明	人口 Logistic 曲线模拟

```
clear
clc
% 读入人口数据(1971-2000)
Y=[33815    33981    34004    34165    34212    34327    34344    34458    34498    34476
34483    34488    34513    34497    34511    34520    34507    34509    34521    34513
34515    34517    34519    34519    34521    34521    34523    34525    34525    34527]
% 读入时间变量数据(t=年份-1970)
T=[1    2    3    4    5    6    7    8    9    10    11    12    13    14    15
16    17    18    19    20    21    22    23    24    25    26    27    28    29    30]
% 线性化处理
for t = 1:30,
   x(t)=exp(-t);
   y(t)=1/Y(t);
end
% 计算,并输出回归系数 B
c=zeros(30,1)+1;
X=[c,x'];
B=inv(X'*X)*X'*y'
for i=1:30,
% 计算回归拟合值
    z(i)=B(1,1)+B(2,1)*x(i);
% 计算离差
    s(i)=y(i)-sum(y)/30;
% 计算误差
    w(i)=z(i)-y(i);
end
% 计算离差平方和 S
S=s*s';
% 回归误差平方和 Q
Q=w*w';
% 计算回归平方和 U
U=S-Q;
% 计算,并输出 F 检验值
F=28*U/Q
% 计算非线性回归模型的拟合值
for j=1:30,
    Y(j)=1/(B(1,1)+B(2,1)*exp(-j));
end
% 输出非线性回归模型的拟合曲线(Logisic 曲线)
plot(T,Y)
```

运行该程序后,输出:

(1) 回归系数 B 及 F 检验值:

```
B =
    1.0e-004 *
    0.2902
    0.0182
F =
   47.8774
```

(2) Logistic 模型拟合曲线(见图 1-8)。

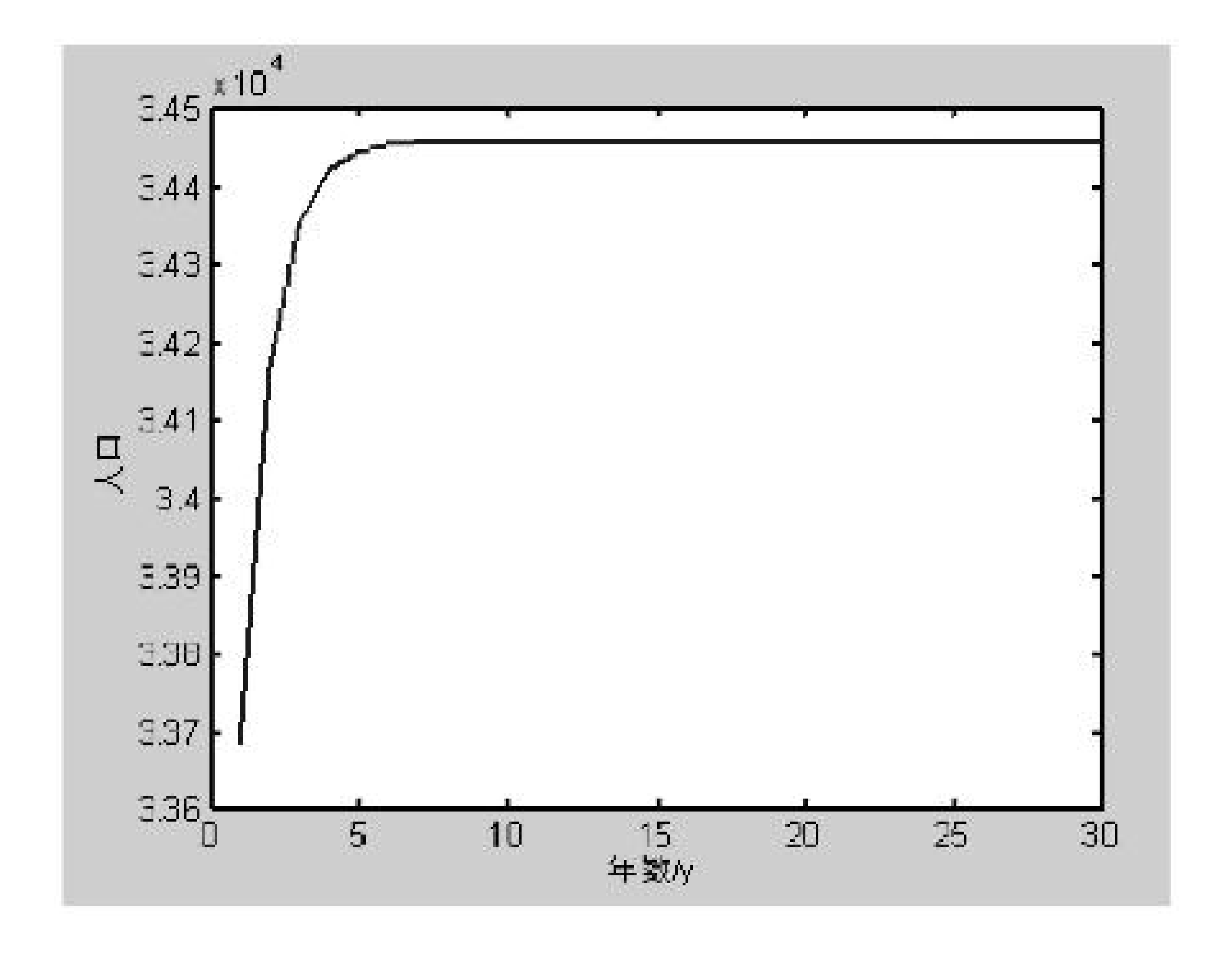


图 1-8　Logistic 模型拟合曲线

1.3.2　薄膜渗透率的测定

1. 问题的提出

某种医用薄膜有允许一种物质的分子穿透它(从高浓度的溶液向低浓度的溶液扩散)的功能,在试制时需测定薄膜被这种分子穿透的能力。测定方法如下:用面积为 S 的薄膜将容器分成体积分别为 V_A、V_B 的两部分(见图 1-9),在两部分中分别注满该物质的两种不同浓度的溶液。此时,该物质分子就会穿过薄膜从高浓度溶液向低浓度溶液扩散。已知通过单位面积薄膜分子扩散的速度与膜两侧溶液的浓度差成正比,比例系数 K 表征了薄膜被该物质分子穿透的能力,称为渗透率。定时测量容器中薄膜某一侧的溶液浓度值,可以确定 K 的数值,试用数学建模的方法解决 K 值的求解问题。

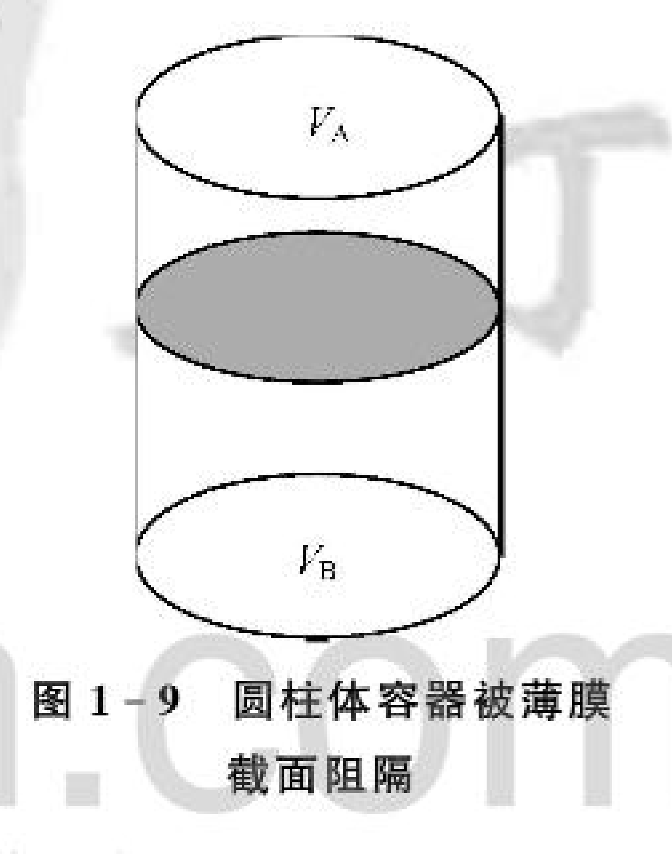


图 1-9　圆柱体容器被薄膜截面阻隔

2. 数学模型的建立

(1) 假　设

① 薄膜两侧的溶液始终是均匀的,即在任何时刻膜两侧的每一处溶液的浓度都是相同的。

② 当两侧浓度不一致时,物质的分子穿透薄膜总是从高浓度溶液向低浓度溶液扩散。

③ 通过单位面积膜分子扩散的速度与膜两侧溶液的浓度差成正比。

④ 薄膜是双向同性的,即物质从膜的任何一侧向另一侧渗透的性能是相同的。

(2) 符号说明

① $C_A(t)$、$C_B(t)$表示 t 时刻膜两侧溶液的浓度。

② a_A、a_B 表示初始时刻膜两侧溶液的浓度(单位:mg/cm^3)。

③ K 表示渗透率。

④ V_A、V_B 表示由薄膜阻隔的容器两侧的体积。

(3) 模型的建立

考察时段$[t,t+\Delta t]$薄膜两侧容器中该物质质量的变化。以容器 A 侧为例，在该时段物质质量增加量为

$$V_A C_A(t+\Delta t)-V_A C_A(t)$$

另一方面由渗透率的定义可知，从 B 侧渗透至 A 侧的该物质的质量为

$$SK(C_B-C_A)\Delta t$$

由质量守恒定律，两者应该相等，于是有

$$V_A C_A(t+\Delta t)-V_A C_A(t)=SK(C_B-C_A)\Delta t$$

两边同除 Δt，令 $\Delta t\to 0$，整理得

$$\frac{dC_A}{dt}=\frac{SK}{V_A}(C_B-C_A) \tag{1-1}$$

且注意到整个容器的溶液中含有该物质的质量应该不变，即有下式成立：

$$V_A C_A(t)+V_B C_B(t)=V_A a_A+V_B a_B \tag{1-2a}$$

$$C_A(t)=a_A+\frac{V_B}{V_A}a_B-\frac{V_B}{V_A}C_B(t) \tag{1-2b}$$

将式(1-2b)代入式(1-1)得

$$\frac{dC_B}{dt}+SK\left(\frac{1}{V_A}+\frac{1}{V_B}\right)C_B=SK\left(\frac{a_A}{V_B}+\frac{a_B}{V_A}\right)$$

再利用初始条件 $C_B(0)=a_B$

解得 $$C_B(t)=\frac{a_A V_A+a_B V_B}{V_A+V_B}+\frac{V_A(a_B-a_A)}{V_A+V_B}e^{-SK\left(\frac{1}{V_A}+\frac{1}{V_B}\right)t}$$

至此，问题归结为利用 C_B 在时刻 t_j 的测量数据 $C_j(j=1,2,\cdots,N)$来辨识参数 K 和 a_A、a_B，对应的数学模型变为求函数：

$$\min E(K,a_A,a_B)=\sum_{j=1}^{n}(C_B(t_j)-C_j)^2$$

令 $$a=\frac{a_A V_A+a_B V_B}{V_A+V_B},\quad b=\frac{V_A(a_B-a_A)}{V_A+V_B}$$

问题转化为求函数

$$E(K,a_A,a_B)=\sum_{j=1}^{n}\left[a+be^{-SK\left(\frac{1}{V_A}+\frac{1}{V_B}\right)t_j}-C_j\right]^2$$

的最小值点(K,a,b)。

3. 求解参数

例如，设 $V_A=V_B=1\,000\ cm^3$，$S=10\ cm^2$，对容器 B 部分溶液浓度的测试结果如表 1-5 所列。

表 1-5 容器 B 部分溶液测试浓度

t_j/s	100	200	300	400	500	600	700	800	900	1000
$C_j/(mg\cdot cm^{-3})$	4.54	4.99	5.35	5.65	5.90	6.10	6.26	6.39	6.50	6.59

此时极小化的函数为

$$E(K,a_A,a_B)=\sum_{j=1}^{10}[a+be^{-0.02Kt_j}-C_j]^2$$

下面用 MATLAB 软件进行计算。

(1) 编写 M 文件(curvefun. m)

```
function f = curvefun(x,tdata)
f = x(1) + x(2) * exp( - 0.02 * x(3) * tdata);        % 其中,x(1) = a,x(2) = b,x(3) = k
```

(2) 编写程序(test1. m)

```
tdata = linspace(100,1000,10);
cdata = 1e-05. * [454 499 535 565 590 610 626 639 650 659];
x0 = [0.2,0.05,0.05];
x = lsqcurvefit ('curvefun',x0,tdata,cdata)
f= curvefun(x,tdata)
plot(tdata,cdata,'o',tdata,f,'r-')
```

(3) 输出结果

```
x =
    0.0063    -0.0034    0.2542
即表示 k = 0.2542,  a = 0.0063,  b = -0.0034
f =
    0.0043    0.0051    0.0056    0.0059    0.0061    0.0062    0.0062    0.0063
0.0063    0.0063
```

曲线的拟合结果如图 1-10 所示,进一步可求得 $a_B=0.004\ \mathrm{mg/cm^3}$,$a_A=0.01\ \mathrm{mg/cm^3}$。

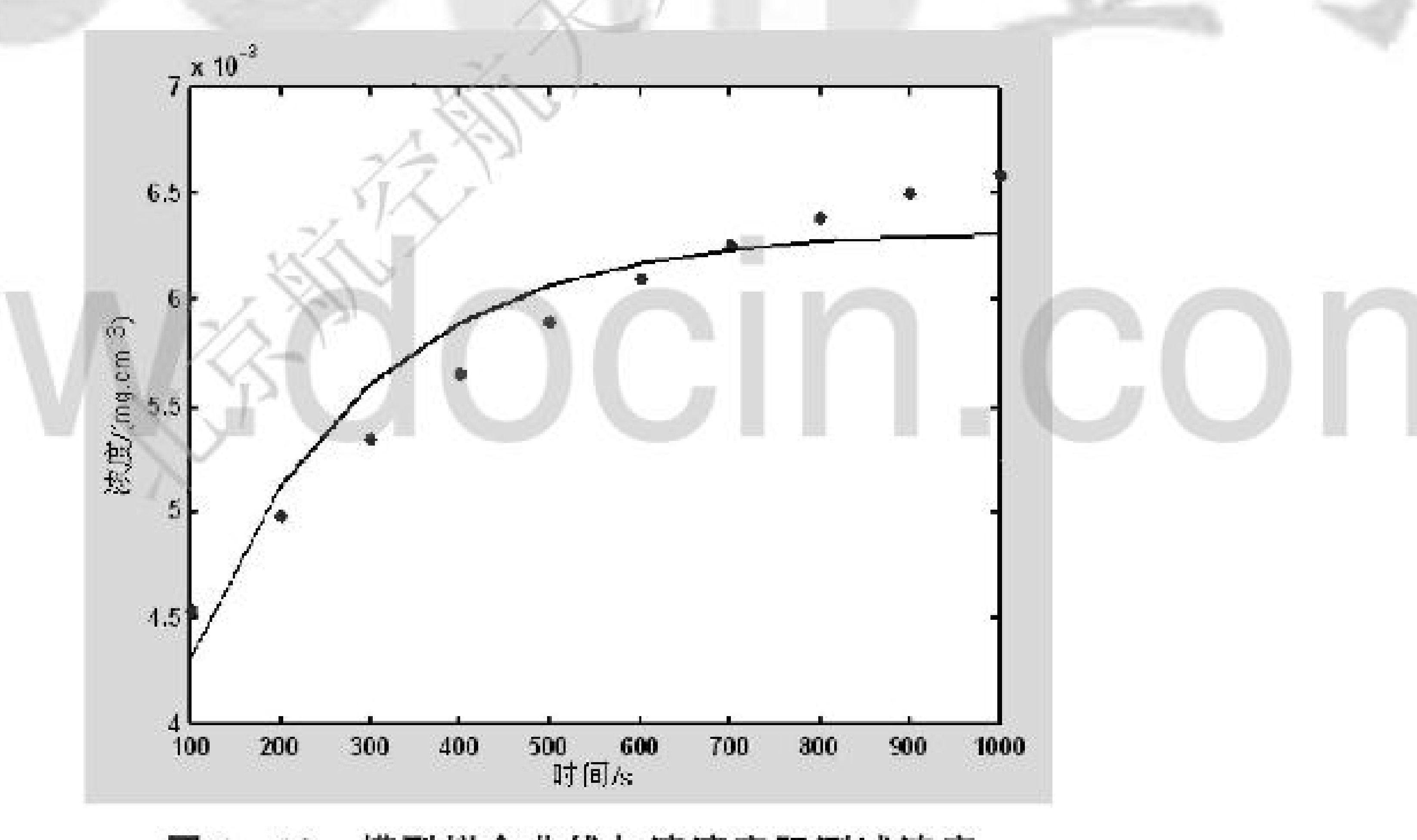


图 1-10 模型拟合曲线与溶液实际测试浓度

1.4 数据的可视化

1.4.1 地形地貌图形的绘制

表 1-6 所列为某次地貌测量所得结果,对一方形区域(x,y 方向均为从 1~10),选择某些地点测量其相对于某水平面高度的数据,要求用这些数据尽量准确地绘制出该地区的地形地貌。

表 1-6　某地区地貌测量结果

	1	2	3	4	5	6	7	8	9	10
1		0.02	−0.12		−2.09		−0.58	−0.08		
2	0.02			−2.38		−4.96				−0.1
3		0.10	1.00		−3.04		−0.53		0.10	
4				3.52						
5	−0.43	−1.98				0.77		2.17		
6			−2.29		0.69		2.59		0.30	
7	−0.09	−0.31				4.27				−0.01
8				5.13	7.40		1.89		0.04	
9	0.1		0.58			1.75		−0.11		
10		−0.01			0.3					0.01

解决此问题的关键是要将未测量地点的高度用数据插值的方法求出来，然后就可以用 MATLAB 绘图方法画出地貌图。

编写以下命令文件，即可解决此问题。

```
[x,y] = meshgrid(1:10);                        % 构造测量网格
h = [0,0.02,-0.12,0,-2.09,0,-0.58,-0.08,0,0;...
0.02,0,0,-2.38,0,-4.96,0,0,0,-0.1;...
0,0.1,1,0,-3.04,0,-0.53,0,0.1,0;...
0,0,0,3.52,0,0,0,0,0,0;...
-0.43,-1.98,0,0,0,0.77,0,2.17,0,0;...
0,0,-2.29,0,0.69,0,2.59,0,0.3,0;...
-0.09,-0.31,0,0,0,4.27,0,0,0,-0.01;...
0,0,0,5.13,7.4,0,1.89,0,0.04,0;...
0.1,0,0.58,0,0,1.75,0,-0.11,0,0;...
0,-0.01,0,0,0.3,0,0,0,0,0.01];                 % 测量数据点
[xi,yi] = meshgrid(1:0.1:10);                  % 构造插值网格
hi = interp2(x,y,h,xi,yi,'spline');            % 二维插值命令
surf(hi);                                      % 画出地貌图
xlable('x'),ylable('y'),zlable('h');           % 添加坐标名称
```

运行此文件，得到如图 1-11 所示的结果，即为该地区的地貌形态图。此处用分段三次样条插值会得到更好的效果。

1.4.2　车灯光源投影区域的绘制(CUMCM 2002 A)

2002 年 CUMCM 的 A 题，其中的一个要求是绘制投影区域。需要先建立车灯投影的数学模型，然后再根据模型绘出投影效果图。这里不再详述建模过程，仅给出一段绘制投影区域的 MATLAB 程序，以说明 MATLAB 绘图技术在建模比赛中的重要性及程序绘图的方法。

利用 MATLAB，根据得到的线光源长度，用投点法可画出测试屏上反射光的亮区，MATLAB程序如下：

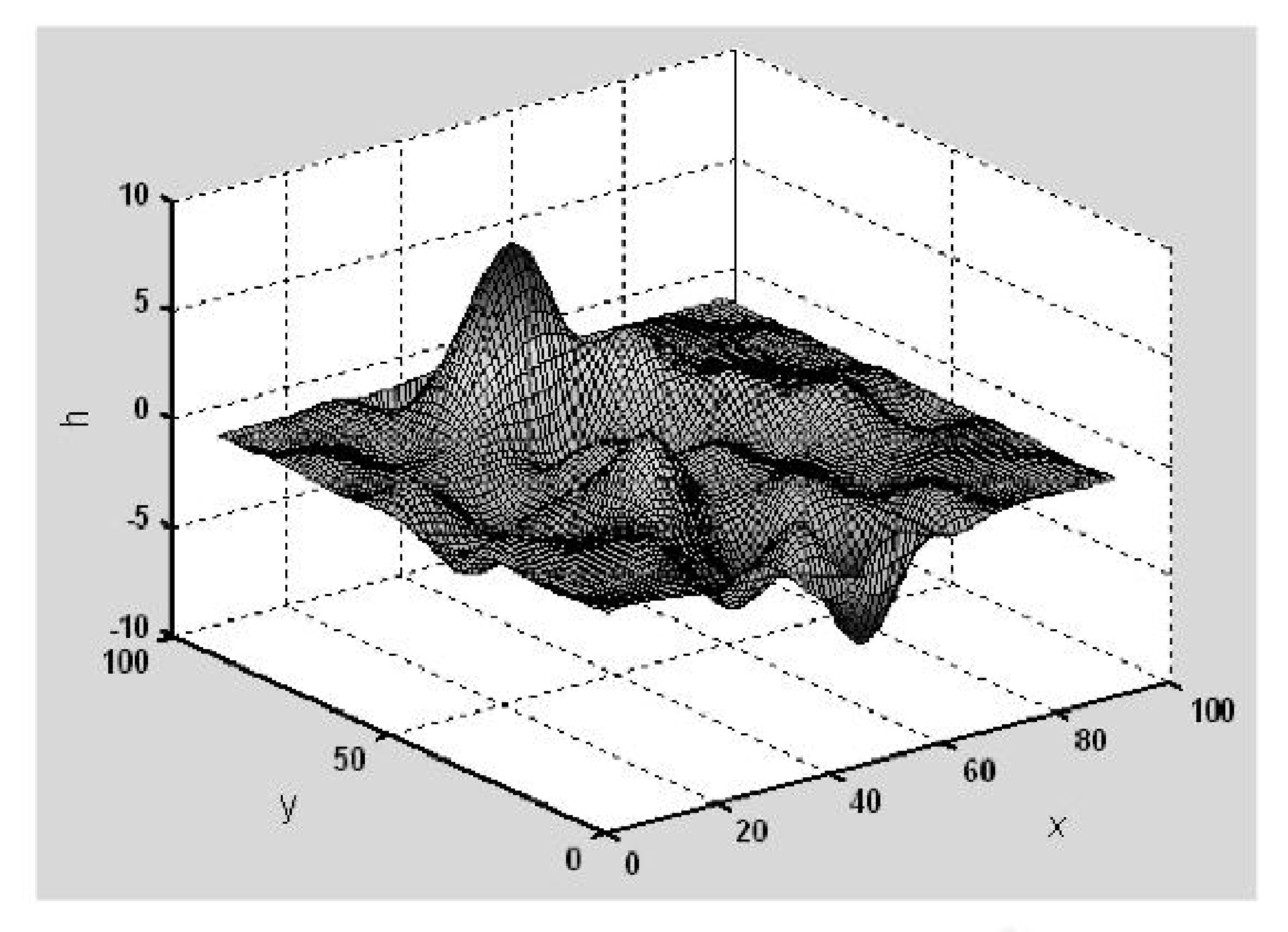


图 1-11　地貌图

```
p = 0.03;x = 25.0216;
for y1 = -0.002:0.0004:0.002
    y0 = (-0.036:0.001:0.036)' * ones(1,73);
    z0 = ones(73,1) * (-0.036:0.001:0.036);
    x0 = (y0.^2 + z0.^2)/(2 * p);
    xn = (p^3 + 4 * x0 * 2 * p. * x0 + p * (-4 * y1 * y0 + 3 * 2 * p * x0))./(2 * (p^2 + 2 * p * x0));
    yn = (2 * p * x0. * y0 + p^2 * (-y1 + y0) + y1 * (y0.^2 - z0.^2))./(p^2 + 2 * p * x0);
    zn = (p^2 + 2 * p * x0 + 2 * y1 * y0). * z0./(p^2 + 2 * p * x0);
    y = y0 + (yn - y0). * (x - x0)./(xn - x0);
    z = z0 + (zn - z0). * (x - x0)./(xn - x0);
    plot(y,z,'b.')
    xlabel('y')
    ylabel('z')
    hold on
end
```


运行程序，即得到反射光在测试屏上的亮区，如图 1-12 所示。

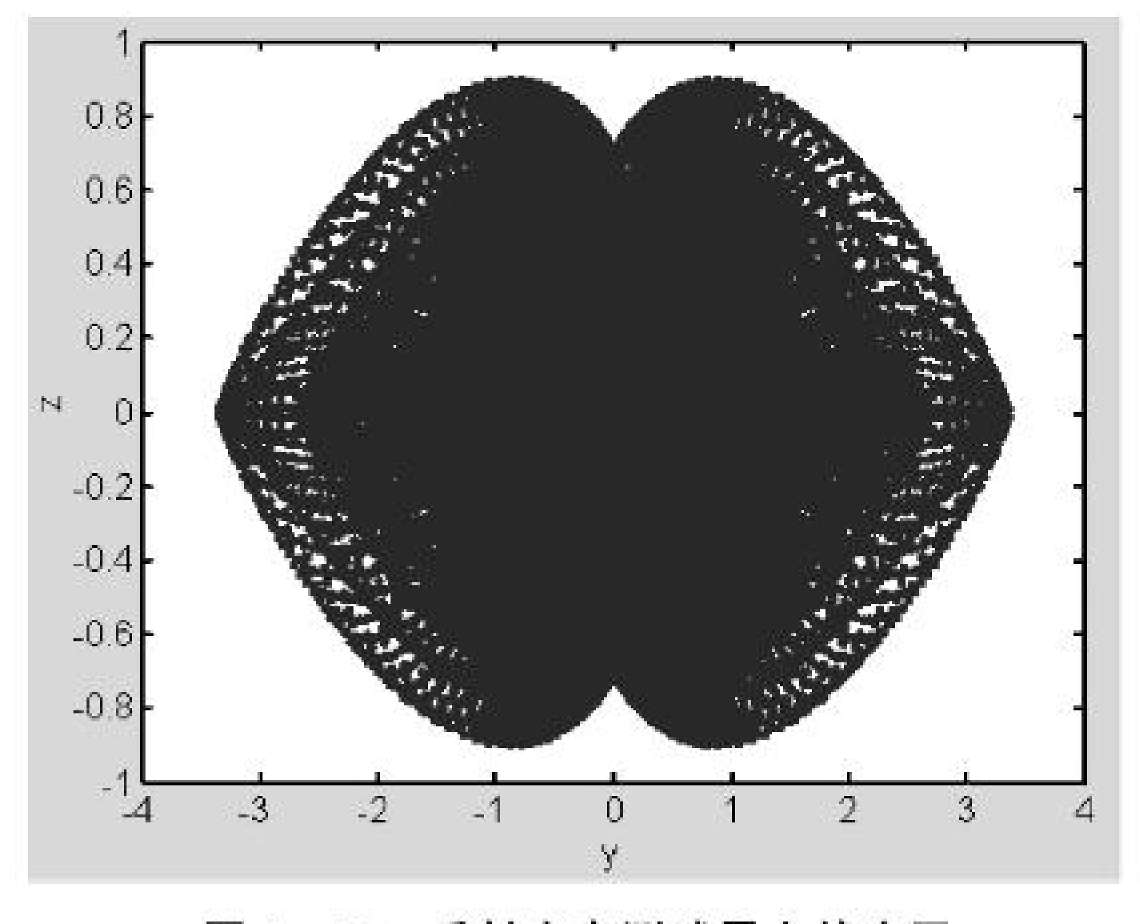


图 1-12　反射光在测试屏上的亮区

第 2 章

规划问题的 MATLAB 求解

规划类问题是常见的数学建模问题,离散系统的优化问题一般都可以通过规划模型来求解。所以,在建模竞赛中,能够快速求解规划类问题是数学建模队员的基本素质。利用 MATLAB提供的强大的规划模型求解命令,可以简单快速地得到所要的结果。一般的标准规划模型都可以用这些命令直接进行求解,本章主要介绍常见规划模型的 MATLAB 求解,包括线性规划、整数规划和非线性规划三个部分。掌握这几部分的操作,可以解决大部分的规划模型的求解问题。

2.1 线性规划

人们在生产实践中,经常会遇到如何利用现有资源安排生产,以取得最大经济效益的问题。此类问题构成了运筹学的一个重要分支——数学规划,而线性规划(linear programming, LP)则是数学规划的一个重要分支。历史上线性规划理论发展的重要进程如下:

1947 年美国数学家 G. B. Dantzig 提出线性规划的一般数学模型和求解线性规划问题的通用方法——单纯形法,为这门学科奠定了基础;

1947 年美国数学家 J. von 诺伊曼提出对偶理论,开创了线性规划的许多新的研究领域,扩大了它的应用范围和解题能力;

1951 年美国经济学家 T. C. 库普曼斯把线性规划应用到经济领域,为此与康托罗维奇一起获 1975 年诺贝尔经济学奖。

目前,线性规划理论趋向完善,应用范围不断延伸,已经渗透到众多领域,特别是在计算机能处理成千上万个约束条件和决策变量的线性规划问题之后,线性规划的适用领域更为广泛了,已成为现代管理中经常采用的基本方法之一。

2.1.1 线性规划的实例与定义

【例 2-1】 中央电视台为改版后的《非常 6+1》栏目播放两套宣传片。其中,宣传片甲播映时间为 3.5 min,广告时间为 30 s,收视观众为 60 万;宣传片乙播映时间为 1 min,广告时间为 1 min,收视观众为 20 万。广告公司规定每周至少有 3.5 min 广告,而电视台每周只能为该栏目宣传片提供不多于 16 min 的节目时间。电视台每周应播映两套宣传片各多少次,才能使得收视观众最多?

分析:建模是解决线性规划问题的极为重要的环节与技术。一个正确数学模型的建立要求建模者熟悉规划问题的生产和管理内容,明确目标要求和错综复杂的约束条件。本题首先将已知数据列成表 2-1 所示的清单。

表 2-1　题意信息清单

类别	播放片甲	播放片乙	节目要求	
片集时间/min	3.5	1		≤16
广告时间/min	0.5	1	≥3.5	
收视观众/万人	60	20		

设电视台每周应播映片甲 x 次，片乙 y 次，总收视观众为 z 万人

目标函数：
$$\max z=60x+20y \tag{2-1}$$

约束条件：
$$\begin{cases}4x+2y\leqslant 16\\0.5x+y\geqslant 3.5\\x,y\in\mathbf{N}\end{cases} \tag{2-2}$$

其中，变量 x，y 称为决策变量，式(2-1)被称为问题的目标函数，式(2-2)中的几个不等式是问题的约束条件，记为 s. t.（即 subject to）。上述即为一规划问题数学模型的三个要素。由于上面的目标函数及约束条件均为线性函数，故被称为线性规划问题。

在解决实际问题时，把问题归结成一个线性规划数学模型是很重要的一步，但往往也是困难的一步，模型建立的是否恰当直接影响到求解。而选取适当的决策变量，是建立有效模型的关键之一。

2.1.2　线性规划的 MATLAB 标准形式

线性规划的目标函数可以是求最大值，也可以是求最小值；约束条件的不等号可以是小于号，也可以是大于号。为了避免形式多样性带来的不便，MATLAB 中规定线性规划的标准形式为

$$\min_{x} \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}\quad \text{s. t.}\quad \boldsymbol{Ax}\leqslant \boldsymbol{b}$$

其中，$\boldsymbol{c}$ 和 $\boldsymbol{x}$ 为 n 维列向量；$\boldsymbol{b}$ 为 m 维列向量；$\boldsymbol{A}$ 为 $m\times n$ 矩阵。

例如线性规划：

$$\max_{x} \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}\quad \text{s. t.}\quad \boldsymbol{Ax}\geqslant \boldsymbol{b}$$

的 MATLAB 标准型为

$$\min_{x} -\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}\quad \text{s. t.}\quad -\boldsymbol{Ax}\leqslant -\boldsymbol{b}$$

2.1.3　线性规划问题解的概念

一般线性规划问题的标准型为

$$\min z=\sum_{j=1}^{n}c_j x_j \tag{2-3}$$

$$\text{s. t.}\quad \sum_{j=1}^{n}a_{ij}x_j\leqslant b_j,\quad i=1,2,\cdots,m \tag{2-4}$$

满足约束条件式(2-4)的解 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$ 称为线性规划问题的可行解，而使目标函数式(2-3)达到最小值的可行解称为最优解。所有可行解构成的集合称为问题的可行域，记为 R。

2.1.4　求解线性规划的 MATLAB 解法

自从 G. B. Dantzig 于 1947 年提出单纯形法，60 多年来，虽有许多变形体已被开发，但却

保持着同样的基本观念。由于有如下结论：若线性规划问题有有限最优解，则一定有某个最优解是可行区域的一个极点。基于此，单纯形法的基本思路是：先找出可行域的一个极点，据一定规则判断其是否最优，否则转换到与之相邻的另一极点，并使目标函数值更优；如此下去，直到找到某一最优解为止。这里不再详细介绍单纯形法，有兴趣的读者可以参看其他有关线性规划的书籍。下面介绍线性规划的 MATLAB 解法。

MATLAB 中线性规划的标准型为

$$\min_{x} \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}$$

基本函数形式为 linprog(c,A,b)，它的返回值是向量 $\boldsymbol{x}$ 的值。还有其他一些函数调用形式（在 MATLAB 指令窗口运行 help linprog 可以看到所有的函数调用形式），如

[x,fval]=linprog(c,A,b,Aeq,beq,LB,UB,X0,OPTIONS)

其中，fval 返回目标函数的值；Aeq 和 beq 对应等式约束 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$；LB 和 UB 分别是变量 x 的下界和上界；$\boldsymbol{X}$0 是 x 的初始值；OPTIONS 是控制参数。

【例 2-2】 求解线性规划问题：

$$\min z = 2x_1 + 3x_2 + x_3$$

$$\text{s.t.} \begin{cases} x_1 + 4x_2 + 2x_3 \geqslant 8 \\ 3x_1 + 2x_2 \geqslant 6 \\ x_1, x_2, x_3 \geqslant 0 \end{cases}$$

解：编写 MATLAB 程序如下：

```
c = [2;3;1];
a = [1,4,2;3,2,0];
b = [8;6];
[x,y] = linprog(c, - a, - b,[],[],zeros(3,1))
```

【例 2-3】 求下面的优化问题：

$$\min z = -5x_1 - 4x_2 - 6x_3$$

$$\text{s.t.} \begin{cases} x_1 - x_2 + x_3 \leqslant 20 \\ 3x_1 + 2x_2 + 4x_3 \leqslant 42 \\ 3x_1 + 2x_2 \leqslant 30 \\ 0 \leqslant x_1, 0 \leqslant x_2, 0 \leqslant x_3 \end{cases}$$

解：

```
>> f = [-5; -4; -6];
>> A =  [1 -1  1;3  2  4;3  2  0];
>> b = [20; 42; 30];
>> lb = zeros(3,1);
>> [x,fval,exitflag,output,lambda] = linprog(f,A,b,[],[],lb)
```

结果为

```
x =                        % 最优解
    0.0000
   15.0000
    3.0000
fval =                     % 最优值
```

```
-78.0000
  exitflag =                    % 收敛
        1
  output =
        iterations: 6           % 迭代次数
      cgiterations: 0
         algorithm: 'lipsol'    % 所使用规则
  lambda =
      ineqlin: [3x1 double]
        eqlin: [0x1 double]
        upper: [3x1 double]
        lower: [3x1 double]
 >> lambda.ineqlin
  ans =
      0.0000
      1.5000
      0.5000
 >> lambda.lower
  ans =
      1.0000
      0.0000
      0.0000
```


【例 2-4】 求解下列线性规划问题：

$$\max z = 2x_1 + 3x_2 - 5x_3$$

$$\text{s. t.} \begin{cases} x_1 + x_2 + x_3 = 7 \\ 2x_1 - 5x_2 + x_3 \geqslant 10 \\ x_1, x_2, x_3 \geqslant 0 \end{cases}$$

解：

(1) 编写 M 文件。


```
c = [2;3; -5];
a = [-2,5, -1]; b = -10;
aeq = [1,1,1];
beq = 7;
% 是求最大值而不是最小值,注意这里是" - c"而不是"c"
x = linprog( - c,a,b,aeq,beq,zeros(3,1))
value = c' * x
```

(2) 将 M 文件存盘,并命名为 example1. m。

(3) 在 MATLAB 指令窗口运行 example1 即可得所求结果。

【例 2-5】 求解下列线性规划问题：

$$\max z = 170.8582x_1 - 17.7254x_2 + 41.2582x_3 + 2.2182x_4 + 131.8182x_5 - 500000$$

$$\text{s. t.}\begin{cases}x_1-0.17037x_2-0.5324x_3+x_5\leqslant 0\\0.17037x_2+0.5324x_3\leqslant 888115\\x_1+32\%x_2+x_3\leqslant 166805\\x_2\leqslant 521265.625\\x_3+x_4\leqslant 683400\\x_4+x_5\geqslant 660000\\x_j\geqslant 0,j=1,2,3,4,5\end{cases}$$

解:为了便于求解,将上述求解最大值线性规划问题转化成求解最小值问题:

$$\min z'=-170.8582x_1+17.7254x_2-41.2582x_3-2.2182x_4-131.8182x_5+500000x_6$$

$$\text{s. t.}\begin{cases}x_1-0.17037x_2-0.5324x_3+x_5\leqslant 0\\0.17037x_2+0.5324x_3\leqslant 888115\\x_1+32\%x_2+x_3\leqslant 166805\\x_2\leqslant 521265.625\\x_3+x_4\leqslant 683400\\-x_4-x_5\leqslant -660000\\x_6=1\\x_j\geqslant 0,j=1,2,3,4,5\end{cases}$$

MATLAB 源程序:

```
f = [-170.8582 17.7254 -41.2582 -2.2182 -131.8182 500000];
A = [1 -0.17037 -0.5324 0 1 0;0 0.17037 0.5324 0 0 0;1 0.32 1 0 0 0;0 1 0 0 0 0;0 0 1 1 0 0;0 0 0 -1 -1 0];
b = [0;888115;166805;521265.625;683400;-660000];
Aeq = [0 0 0 0 0 1];
beq = [1];
lb = [0;0;0;0;0;0];
[x,fval,exitflag,output,lambda] = linprog(f,A,b,Aeq,beq,lb,[])
```

程序输出结果为

```
x =
  1.0e+005 *
   0.00000000000000
   1.70617739889132
   1.12207323235472
   5.71192676764526
   0.88807323235476
   0.00001000000000
fval = -1.407864558820066e+007
```

即 $x_1=0,\quad x_2=170618,\quad x_3=112207,\quad x_4=571193,\quad x_5=88807,\quad x_6=1$

$$\min z'=-14078646,\quad \max z=-\min z'=-(-14078646)=14078646$$

2.2 非线性规划

2.2.1 非线性规划的实例与定义

如果目标函数或约束条件中包含非线性函数，就称这种规划问题为非线性规划问题。一般说来，解非线性规划要比解线性规划问题困难得多，而且非线性规划也不像线性规划那样有单纯形法这一通用方法。非线性规划目前还没有适于各种问题的一般算法，各种方法都有自己特定的适用范围。

下面通过实例归纳出非线性规划数学模型的一般形式，介绍有关非线性规划的基本概念。

【例 2-6】 投资决策问题：某企业有 n 个项目可供选择投资，并且至少要对其中一个项目投资。已知该企业拥有总资金 A 元，投资于第 $i(i=1,2,\cdots,n)$ 个项目需花资金 a_i 元，并预计可收益 b_i 元。试选择最佳投资方案。

解：设投资决策变量为：

$$x_i = \begin{cases} 1, & \text{决定投资第 } i \text{ 个项目} \\ 0, & \text{决定不投资第 } i \text{ 个项目} \end{cases}, \quad i = 1,2,\cdots,n$$

则投资总额为 $\sum_{i=1}^{n} a_i x_i$，投资总收益为 $\sum_{i=1}^{n} b_i x_i$。

因为该公司至少要对一个项目投资，并且总的投资金额不能超过总资金 A，故有限制条件：

$$0 < \sum_{i=1}^{n} a_i x_i \leqslant A$$

由于 $x_i(i=1,2,\cdots,n)$ 只取值 0 或 1，所以还有

$$x_i(1-x_i) = 0, \quad i = 1,2,\cdots,n$$

另外，该公司至少要对一个项目投资，因此有

$$\sum_{i=1}^{n} x_i \geqslant 1$$

最佳投资方案应是投资额最小而总收益最大的方案，所以这个最佳投资决策问题归结为在总资金以及决策变量（取 0 或 1）的限制条件下，极大化利润（即总收益）与总投资之差。因此，其数学模型为

$$\max Q = \sum_{i=1}^{n} b_i x_i - \sum_{i=1}^{n} a_i x_i$$

$$\text{s.t.} \begin{cases} 0 < \sum_{i=1}^{n} a_i x_i \leqslant A \\ x_i(1-x_i) = 0, \quad i = 1,2,\cdots,n \\ \sum_{i=1}^{n} x_i \geqslant 1 \end{cases}$$

例 2-6 是在一组等式或不等式的约束下，求一个函数的最大值（或最小值）问题，其中目标函数或约束条件中至少有一个非线性函数，这类问题称之为非线性规划问题，简记为（NP）。可概括为一般形式：

$$\min f(\boldsymbol{x})$$

$$\text{s.t.}\quad h_j(x) \leqslant 0,\quad j=1,2,\cdots,q$$
$$g_i(x)=0,\quad i=1,2,\cdots,p$$

其中,$\boldsymbol{x}=[x_1 \quad x_2 \quad \cdots \quad x_n]^{\mathrm{T}}$ 称为模型(NP)的决策变量;$\boldsymbol{f}$ 称为目标函数;$g_i(i=1,2,\cdots,p)$和 $h_i(j=1,2,\cdots,q)$称为约束函数。另外,$g_i(x)=0(i=1,2,\cdots,p)$称为等式约束;$h_j(x)\leqslant 0(j=1,2,\cdots,q)$称为不等式约束。

2.2.2 非线性规划的 MATLAB 解法

在 MATLAB 中非线性规划的数学模型可写成以下形式:

$$\min f(x)$$
$$\text{s.t.}\begin{cases}\boldsymbol{Ax}\leqslant \boldsymbol{B}\\ \text{Aeq}\cdot\boldsymbol{x}=\text{Beq}\\ \boldsymbol{C}(\boldsymbol{x})\leqslant 0\\ \text{Ceq}(\boldsymbol{x})=0\end{cases}$$

其中,$f(x)$是标量函数;$\boldsymbol{A}$,$\boldsymbol{B}$,Aeq,Beq 是相应维数的矩阵和向量;$\boldsymbol{C}(\boldsymbol{x})$,Ceq($\boldsymbol{x}$)是非线性向量函数。

MATLAB 中的命令是

```
X = FMINCON(FUN,X0,A,B,Aeq,Beq,LB,UB,NONLCON,OPTIONS)
```

它的返回值是向量 $\boldsymbol{x}$。其中,FUN 是用 M 文件定义的函数 $f(x)$。X0 是 $\boldsymbol{x}$ 的初始值。A,B,Aeq,Beq 定义了线性约束 $\boldsymbol{Ax}\leqslant\boldsymbol{B}$,Aeq · $\boldsymbol{x}$=Beq,如果没有线性约束,则 A=[],B=[],Aeq=[],Beq=[]。LB 和 UB 是变量 $\boldsymbol{x}$ 的下界和上界,如果上界和下界没有约束,则 LB=[],UB=[];如果 $\boldsymbol{x}$ 无下界,则 LB=-inf;如果 $\boldsymbol{x}$ 无上界,则 UB=inf。NONLCON 是用 M 文件定义的非线性向量函数 $\boldsymbol{C}(\boldsymbol{x})$,Ceq($\boldsymbol{x}$)。OPTIONS 定义了优化参数,可以使用 MATLAB 默认的参数设置。

【例 2-7】 求下列非线性规划问题:

$$\min f(x)=x_1^2+x_2^2+8$$
$$\text{s.t.}\begin{cases}x_1^2-x_2\geqslant 0\\ -x_1-x_2^2+2=0\\ x_1,x_2\geqslant 0\end{cases}$$

(1)编写 M 文件 fun1.m 和 M 文件 fun2.m。

```
function f = fun1(x);
f = x(1)^2 + x(2)^2 + 8;
```

```
function [g,h] = fun2(x);
g = - x(1)^2 + x(2);
h = - x(1) - x(2)^2 + 2;  % 等式约束
```

(2)在 MATLAB 的命令窗口依次输入如下语句:

```
options = optimset;
[x,y] = fmincon('fun1',rand(2,1),[],[],[],[],zeros(2,1),[], ...
'fun2', options)
```

就可以求得当 $x_1=1,x_2=1$ 时，最小值 $y=10$。

【例 2－8】 求下列非线性规划问题：

$$\max z = \sqrt{x_1}+\sqrt{x_2}+\sqrt{x_3}+\sqrt{x_4}$$

$$\text{s. t.}\begin{cases} x_1 \leqslant 400 \\ 1.1x_1+x_2 \leqslant 440 \\ 1.21x_1+1.1x_2+x_3 \leqslant 484 \\ 1.331x_1+1.21x_2+1.1x_3+x_4 \leqslant 532.4 \\ x_i \geqslant 0, \quad i=1,2,3,4 \end{cases}$$

(1)编写 M 文件，定义目标函数：

```
function f = fun44(x)
f = -(sqrt(x(1)) + sqrt(x(2)) + sqrt(x(3)) + sqrt(x(4)));
```

(2)编写 M 文件，定义约束条件：

```
function [g,ceq] = mycon1(x)
g(1) = x(1) - 400;
g(2) = 1.1 * x(1) + x(2) - 440;
g(3) = 1.21 * x(1) + 1.1 * x(2) + x(3) - 484;
g(4) = 1.331 * x(1) + 1.21 * x(2) + 1.1 * x(3) + x(4) - 532.4;
ceq = 0;
```

(3) 编写主程序(既可以编写 M 文件，也可以在 Command Window 直接输入命令)：

```
x0 = [1;1;1;1];lb = [0;0;0;0];ub = [];A = [];b = [];Aeq = [];beq = [];
[x,fval] = fmincon('fun44',x0,A,b,Aeq,beq,lb,ub,'mycon1')
```

程序输出结果为

```
x =
   1.0e + 002  *
   0.84243824470856
   1.07635203745600
   1.28903186524063
   1.48239367919807
fval =
  - 43.08209516098581
```

所以最终结果为

$$\begin{cases} x_1 = 84.24, \quad x_2 = 107.63, \quad x_3 = 128.90, \quad x_4 = 148.23 \\ z = -(-43.08) = 43.08 \end{cases}$$

2.2.3 二次规划

若某非线性规划的目标函数为自变量 x 的二次函数，约束条件又全是线性的，则称这种规划为二次规划。

在 MATLAB 中二次规划的数学模型可表述如下：

$$\min \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x}+\boldsymbol{f}^{\mathrm{T}}\boldsymbol{x}, \quad \text{s. t.} \quad \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}$$

其中，$\boldsymbol{f}$，$\boldsymbol{b}$ 是列向量；$\boldsymbol{A}$ 是相应维数的矩阵；$\boldsymbol{H}$ 是实对称矩阵。实对称矩阵的定义：如果有 n 阶矩阵 $\boldsymbol{A}$，其各个元素都是实数，且满足 $a_{ij}=a_{ji}$（转置为其本身），则称 $\boldsymbol{A}$ 为实对称矩阵。

MATLAB 中求解二次规划的命令是

```
[X,FVAL] = QUADPROG(H,f,A,b,Aeq,beq,LB,UB,X0,OPTIONS)
```

X 的返回值是向量 $\boldsymbol{x}$；FVAL 的返回值是目标函数在 $\boldsymbol{x}$ 处的值。具体细节可以参看在 MATLAB 指令中运行 help quadprog 后的帮助。

【例 2－9】 求解二次规划：

$$\min f(\boldsymbol{x}) = 2x_1^2 - 4x_1x_2 + 4x_2^2 - 6x_1 - 3x_2$$

$$\text{s. t.} \begin{cases} x_1 + x_2 \leqslant 3 \\ 4x_1 + x_2 \leqslant 9 \\ x_1, x_2 \geqslant 0 \end{cases}$$

解：编写如下程序：

```
h = [4, - 4; - 4,8];
f = [ - 6; - 3];
a = [1,1;4,1];
b = [3;9];
[x,value] = quadprog(h,f,a,b,[],[],zeros(2,1))
```

求得

$$x = \begin{pmatrix} 1.9500 \\ 1.0500 \end{pmatrix}, \qquad \min f(\boldsymbol{x}) = -11.0250$$

利用罚函数法，可将非线性规划问题的求解转化为求解一系列无约束极值问题，因而也称这种方法为序列无约束最小化技术，简记为 SUMT（sequential unconstrained minimization technique）。

罚函数法求解非线性规划问题的思想是，利用问题中的约束函数做出适当的罚函数，由此构造出带参数的增广目标函数，把问题转化为无约束非线性规划问题。罚函数主要有两种形式，一种称外罚函数法；另一种称内罚函数法。下面主要介绍外罚函数法。

考虑如下问题：

$$\min f(\boldsymbol{x})$$

$$\text{s. t.} \begin{cases} g_i(x) \leqslant 0, & i = 1,2,\cdots,r \\ h_i(x) \geqslant 0, & i = 1,2,\cdots,s \\ k_i(x) = 0, & i = 1,2,\cdots,t \end{cases}$$

取一个充分大的数 $M>0$，构造函数

$$\boldsymbol{P}(\boldsymbol{x},\boldsymbol{M}) = \boldsymbol{f}(\boldsymbol{x}) + M\sum_{i=1}^{r}\max(g_i(x),0) - M\sum_{i=1}^{s}\min(h_i(x),0) + M\sum_{i=1}^{t}|k_i(x)|$$

或 $$\boldsymbol{P}(\boldsymbol{x},\boldsymbol{M}) = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{M}_1\max(\boldsymbol{G}(\boldsymbol{x}),0) + \boldsymbol{M}_2\min(\boldsymbol{H}(\boldsymbol{x}),0) + \boldsymbol{M}_3\|\boldsymbol{K}(\boldsymbol{x})\|$$

其中，

$$G(x)=\begin{bmatrix}g_1(x)\\g_2(x)\\\vdots\\g_r(x)\end{bmatrix},\quad H(x)=\begin{bmatrix}h_1(x)\\h_2(x)\\\vdots\\h_s(x)\end{bmatrix},\quad K(x)=\begin{bmatrix}k_1(x)\\k_2(x)\\\vdots\\k_t(x)\end{bmatrix}$$

M_1, M_2, M_3 为适当的行向量，MATLAB 中可以直接利用 max 和 min 函数。则以增广目标函数 $P(x,M)$ 为目标函数的无约束极值问题

$$\min P(x,M)$$

的最优解 x 也是原问题的最优解。

【例 2-10】 求下列非线性规划：

$$\min f(x)=x_1^2+x_2^2+8$$

$$\text{s.t.}\begin{cases}x_1^2-x_2\geqslant 0\\-x_1-x_2^2+2=0\\x_1,x_2\geqslant 0.\end{cases}$$

解：

(1) 编写 M 文件 test.m。

```
function g = test(x);
M = 50000;
f = x(1)^2 + x(2)^2 + 8;
g = f - M * min(x(1),0) - M * min(x(2),0) - M * min(x(1)^2 - x(2),0)...
   + M * abs( - x(1) - x(2)^2 + 2);
```

(2) 在 MATLAB 命令窗口输入：

```
[x,y] = fminunc('test',rand(2,1))
```

即可求得问题的解。

2.3 整数规划

2.3.1 整数规划的定义

规划中的变量（部分或全部）限制为整数时，称为整数规划。若在线性规划模型中，变量限制为整数，则称为整数线性规划。目前所流行的求解整数规划的方法，往往只适用于整数线性规划。目前还没有一种方法能有效地求解一切整数规划。常见的整数规划问题的求解算法有以下几种：

(1) 分枝定界法：可求纯或混合整数线性规划。

(2) 割平面法：可求纯或混合整数线性规划。

(3) 隐枚举法：用于求解 0-1 整数规划，有过滤隐枚举法和分枝隐枚举法。

(4) 匈牙利法：解决指派问题（0-1 规划特殊情形）。

(5) 蒙特卡罗法：求解各种类型规划。

2.3.2 0-1 整数规划

0-1 整数规划是整数规划中的特殊情形,它的变量 x_j 仅取值 0 或 1,这时 x_j 称为 0-1 变量,或称二进制变量。x_j 仅取值 0 或 1,这个条件可由下述约束条件表示:$0 \leqslant x_j \leqslant 1, x_j \in N$ 或 $x_i(1-x_i)=0, i=1,2,\cdots,n$。在实际问题中,如果引入 0-1 变量,就可以把有各种情况需要分别讨论的线性规划问题统一在一个问题中讨论了。

下面举例说明一种解 0-1 整数规划的隐枚举法。所谓隐枚举法是指在枚举过程中引入过滤条件,而与之对应的是显枚举法(一种穷举法)。

【例 2-11】

$$\max z = 3x_1 - 2x_2 + 5x_3$$

$$\text{s.t.}\begin{cases} x_1 + 2x_2 - x_3 \leqslant 2 \\ x_1 + 4x_2 + x_3 \leqslant 4 \\ x_1 + x_2 \leqslant 3 \\ 4x_2 + x_3 \leqslant 6 \\ x_1, x_2, x_3 = 0 \text{ 或 } 1 \end{cases}$$

求解思路及改进措施:

(1) 先试探性地求一个可行解,易看出 $(x_1, x_2, x_3)=(1,0,0)$ 满足约束条件,故为一个可行解,且相应的目标函数值为 $z=3$。

(2) 因为是求极大值问题,故求最优解时,凡是目标值 $z<3$ 的解不必检验是否满足约束条件即可删除,因它肯定不是最优解,于是应增加一个约束条件(目标值下界):

$$3x_1 - 2x_2 + 5x_3 \geqslant 3$$

称该条件为过滤条件(filtering contraint)。从而原问题等价于:

$$\max z = 3x_1 - 2x_2 + 5x_3$$

$$\text{s.t.}\begin{cases} 3x_1 - 2x_2 + 5x_3 \geqslant 3 & (2-5\text{a}) \\ x_1 + 2x_2 - x_3 \leqslant 2 & (2-5\text{b}) \\ x_1 + 4x_2 + x_3 \leqslant 4 & (2-5\text{c}) \\ x_1 + x_2 \leqslant 3 & (2-5\text{d}) \\ 4x_2 + x_3 \leqslant 6 & (2-5\text{e}) \\ x_1, x_2, x_3 = 0 \text{ 或 } 1 & (2-5\text{f}) \end{cases}$$

若用全部枚举法,3 个变量共有 8 种可能的组合,依次检验这 8 种组合,看其是否满足条件式(2-5a)~式(2-5e),对某个组合,若它不满足式(2-5a),即不满足过滤条件,则式(2-5b)~式(2-5e)即可行性条件不必再检验;若它满足式(2-5a)~式(2-5e)且相应的目标值严格大于 3,则进行步骤(3)。

(3) 改进过滤条件。

(4) 由于对每个组合首先计算目标值以验证过滤条件,故应优先计算目标值 z,这样可提前抬高过滤门槛,以减少计算量。

按上述思路与方法,例 2-11 的求解过程可由表 2-2 表示。

表 2-2 例 2-11 的求解过程

(x_1,x_2,x_3)	目标值	约束条件					过滤条件
		式(2-5a)	式(2-5b)	式(2-5c)	式(2-5d)	式(2-5e)	
(0,0,0)	0	×					
(1,0,0)	3	√	√	√	√	√	$3x_1-2x_2+5_3\geqslant 3$
(0,1,0)	−2	×					
(0,0,1)	5	√	√	√	√	√	$3x_1-2x_2+5_3\geqslant 5$
(1,1,0)	1	×					
(1,0,1)	8	√	√	√	√	√	$3x_1-2x_2+5_3\geqslant 8$
(1,1,1)	6	×					
(0,1,1)	3	×					

从而得最优解$(x_1^*,x_2^*,x_3^*)=(1,0,1)$，最优值 $z^*=8$。

2.3.3 随机取样计算法

前面介绍的常用的整数规划求解方法，主要是针对线性整数规划而言，而对于非线性整数规划目前尚未有一种成熟而有效的求解方法，因为非线性规划本身的通用有效解法尚未找到，更何况是非线性整数规划。

然而，尽管整数规划由于限制变量为整数而增加了难度，然而又由于整数解是有限个，于是为枚举法提供了方便。当然，在自变量维数很大和取值范围很宽的情况下，企图用显枚举法(即穷举法)计算出最优值是不现实的，但是应用概率理论可以证明，在有一定计算量的情况下，完全可以得出一个满意解。

【例 2-12】 已知非线性整数规划为

$$\max z=x_1^2+x_2^2+3x_3^2+4x_4^2+2x_5^2-8x_1-2x_2-3x_3-x_4-2x_5$$

$$\text{s.t.}\begin{cases}0\leqslant x_i\leqslant 99, & i=1,2,\cdots,5\\ x_1+x_2+x_3+x_4+x_5\leqslant 400\\ x_1+2x_2+2x_3+x_4+6x_5\leqslant 800\\ 2x_1+x_2+6x_3\leqslant 200\\ x_3+x_4+5x_5\leqslant 200\end{cases}$$

解：(1)首先编写 M 文件 mente.m，定义目标函数 $\boldsymbol{f}$ 和约束向量函数 $\boldsymbol{g}$，程序如下：

```
function [f,g] = mengte(x);
f = x(1)^2 + x(2)^2 + 3 * x(3)^2 + 4 * x(4)^2 + 2 * x(5) - 8 * x(1) - 2 * x(2) - 3 * x(3) - x(4) - 2 * x(5);
g(1) = sum(x) - 400;
g(2) = x(1) + 2 * x(2) + 2 * x(3) + x(4) + 6 * x(5) - 800;
g(3) = 2 * x(1) + x(2) + 6 * x(3) - 200;
g(4) = x(3) + x(4) + 5 * x(5) - 200;
```

(2) 编写如下程序求问题的解：

```
rand('state',sum(clock));
p0 = 0;
```

```
tic
for i = 1:10^5
    x = 99 * rand(5,1);
x1 = floor(x);x2 = ceil(x);
[f,g] = mengte(x1);
if sum(g <= 0) == 4
    if p0 <= f
        x0 = x1;p0 = f;
    end
end
[f,g] = mengte(x2);
if sum(g <= 0) == 4
    if p0 <= f
        x0 = x2;p0 = f;
    end
end
end
x0,p0
toc
```

运行程序，即得到如下的输出结果：

```
ans =
     44     94      3     97     11
p0 =
        47789
```

需要说明的是，由于是随机取样，所以每次运行的结果都可能不一致，但误差不会太大。关于整数规划模型的求解，虽然 MATLAB 可以解决，但对于比较标准的整数规划模型，建议还是首先考虑用 Lingo 求解，这方面 Lingo 比较有优势，实在不行，再考虑用 MATLAB 来解决，MATLAB 的通用性比较强。

第 3 章

灰色预测及其 MATLAB 实现

在数学建模中经常会遇到数据的预测问题,有些赛题中,数据的预测占主导地位,如表 3-1 所列。

表 3-1　历届 CUMCM 数据预测题目

年　度	类　别	题　目	命题人
2003	A 题	SARS 的传播问题	CUMCM 组委会
2005	A 题	长江水质的评价和预测问题	韩中庚
2006	B 题	艾滋病疗法的评价及疗效的预测问题	边馥萍
2007	A 题	中国人口增长预测问题	唐云

有些问题则是需要在求解的过程中进行数据预测,如 2009 年 CUMCM D 题"会议筹备"对与会人数的确定等。灰色模型(gray model,又称灰色理论)有严格的理论基础,最大优点是实用。用灰色模型预测的结果比较稳定,不仅适用于大数据量的预测,在数据量较少时(数据只要多于 3 个即可)预测结果依然较准确。

3.1 灰色预测基础知识

灰色系统理论认为:系统的行为现象尽管是朦胧的,数据是复杂的,但它毕竟是有序的,是有整体功能的。在建立灰色预测模型之前,需先对原始时间序列进行数据处理,经过数据预处理后的数据序列称为生成列。对原始数据进行预处理,不是寻求它的统计规律和概率分布,而是将杂乱无章的原始数据列通过一定的方法处理,变成有规律的时间序列数据,即以数找数的规律,再建立动态模型。灰色系统常用的数据处理方式有累加和累减两种,通常用累加方法。

灰色预测通过鉴别系统因素之间发展趋势的相异程度,并对原始数据进行生成处理来寻找系统变动的规律,生成有较强规律性的数据序列,然后建立相应的微分方程模型,从而预测事物的未来发展趋势。灰色预测的数据是通过生成数据的模型所得到的预测值的逆处理结果。灰色预测是以灰色模型为基础的,在诸多的灰色模型中,以灰色系统中单序列一阶线性微分方程模型 GM(1,1)模型最为常用。下面简要地介绍 GM(1,1)模型。

设有原始数据列 $\boldsymbol{x}^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$,$n$ 为数据个数。

如果根据 $\boldsymbol{x}^{(0)}$数据列建立 GM(1,1)来实现预测功能,则基本步骤如下:

(1) 原始数据累加以便弱化随机序列的波动性和随机性,得到新数据序列:

$$\boldsymbol{x}^{(1)}=(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n))$$

其中,$x^{(1)}(t)$中各数据表示对应前几项数据的累加。

$$x^{(1)}(t)=\sum_{k=1}^{t}x^{(0)}(k),\qquad t=1,2,\cdots,n$$

或 $$x^{(1)}(t+1)=\sum_{k=1}^{t+1}x^{(0)}(k),\qquad t=1,2,\cdots,n$$

(2) 对 $x^{(1)}(t)$ 建立 $x^{(1)}(t)$ 的一阶线性微分方程：

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}+ax^{(1)}=u$$

其中，a，u 为待定系数，分别称为发展系数和灰色作用量，a 的有效区间是 $(-2,\ 2)$，并记 a，u 构成的矩阵为 $\hat{\boldsymbol{a}}=\begin{pmatrix}a\\u\end{pmatrix}$。只要求出参数 a，u，就能求出 $x^{(1)}(t)$，进而求出 $\boldsymbol{x}^{(0)}$ 的未来预测值。

(3) 对累加生成数据做均值生成 $\boldsymbol{B}$ 与常数项向量 $\boldsymbol{Y}_n$，即

$$\boldsymbol{B}=\begin{bmatrix}0.5(x^{(1)}(1)+x^{(1)}(2))\\0.5(x^{(1)}(2)+x^{(1)}(3))\\0.5(x^{(1)}(n-1)+x^{(1)}(n))\end{bmatrix},\qquad \boldsymbol{Y}_n=(x^{(0)}(2),x^{(0)}(3),\cdots,x^{(0)}(n))^{\mathrm{T}}$$

(4) 用最小二乘法求解灰参数 $\hat{\boldsymbol{a}}$，则

$$\hat{\boldsymbol{a}}=\begin{pmatrix}a\\u\end{pmatrix}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}_n$$

(5) 将灰参数 $\hat{\boldsymbol{a}}$ 代入 $\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}+ax^{(1)}=u$，并对 $\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}+ax^{(1)}=u$ 进行求解，得

$$\hat{x}^{(1)}(t+1)=\left(x^{(0)}(1)-\frac{u}{a}\right)e^{-at}+\frac{u}{a}$$

由于 $\hat{\boldsymbol{a}}$ 是通过最小二乘法求出的近似值，所以 $\hat{x}^{(1)}(t+1)$ 是一个近似表达式，为了与原序列 $x^{(1)}(t+1)$ 区分开来，故记为 $\hat{x}^{(1)}(t+1)$。

(6) 对函数表达式 $\hat{x}^{(1)}(t+1)$ 及 $\hat{x}^{(1)}(t)$ 进行离散，并将二者做差以便还原 $\boldsymbol{x}^{(0)}$ 原序列，得到近似数据序列 $\hat{x}^{(0)}(t+1)$ 如下：

$$\hat{x}^{(0)}(t+1)=\hat{x}^{(1)}(t+1)-\hat{x}^{(1)}(t)$$

(7) 对建立的灰色模型进行检验，步骤如下：

① 计算 $x^{(0)}$ 与 $\hat{x}^{(0)}(t)$ 之间的残差 $e^{(0)}(t)$ 和相对误差 $q(x)$：

$$e^{(0)}(t)=x^{(0)}-\hat{x}^{(0)}(t)$$

$$q(x)=\frac{e^{(0)}(t)}{x^{(0)}(t)}$$

② 求原始数据 $\boldsymbol{x}^{(0)}$ 的均值以及方差 s_1。

③ 求 $e^{(0)}(t)$ 的平均值 $\bar{q}$ 以及残差的方差 s_2。

④ 计算方差比 $C=\frac{s_2}{s_1}$。

⑤ 求小误差概率 $P=P\{|e(t)|<0.6745s_1\}$。

⑥ 灰色模型精度检验如表 3-2 所列。

表 3-2　灰色模型精度检验对照表

等　级	相对误差 q	方差比 C	小误差概率 P
Ⅰ级	<0.01	<0.35	>0.95
Ⅱ级	<0.05	<0.50	<0.80

续表 3-2

等　级	相对误差 q	方差比 C	小误差概率 P
Ⅲ级	<0.10	<0.65	<0.70
Ⅳ级	>0.20	>0.80	<0.60

在实际应用过程中，检验模型精度的方法并不唯一。可以利用上述方法进行模型的检验，也可以根据 $q(x)$ 的误差百分比并结合预测数据与实际数据之间的测试结果酌情认定模型是否合理。

(8) 利用模型进行预测：

$$\hat{x}^{(0)} = \left[\underbrace{\hat{x}^{(0)}(1),\hat{x}^{(0)}(2),\cdots,\hat{x}^{(0)}(n)}_{\text{原数列的模拟}},\underbrace{\hat{x}^{(0)}(n+1),\cdots,\hat{x}^{(0)}(n+m)}_{\text{未来数列的预测}}\right]$$

3.2 灰色预测的 MATLAB 程序

3.2.1 典型程序结构

灰色预测中有很多关于矩阵的运算，这可是 MATLAB 的特长，所以用 MATLAB 是实现灰色预测过程的首选。用 MATLAB 编写灰色预测程序时，可以完全按照预测模型的求解步骤，即

(1) 对原始数据进行累加。

(2) 构造累加矩阵 $\boldsymbol{B}$ 与常数向量。

(3) 求解灰参数。

(4) 将参数带入预测模型进行数据预测。

下面以某公司收入预测问题为例介绍灰色预测的 MATLAB 实现过程。

已知某公司 1999—2008 年的利润为(单位：元/年)：[89 677，99 215，109 655，120 333，135 823，159 878，182 321，209 407，246 619，300 670]，现在要预测该公司未来几年的利润情况。

具体的 MATLAB 程序见 P3-1。

程序编号	P3-1	文件名称	main0301.m	说明	灰色预测公司的利润

```
clear
syms a b;
c = [a b];
A = [89677,99215,109655,120333,135823,159878,182321,209407,246619,300670];
B = cumsum(A);                         % 原始数据累加
n = length(A);
for i = 1:(n-1)
    C(i) = (B(i) + B(i+1))/2;          % 生成累加矩阵
end
% 计算待定参数的值
D = A;D(1) = [];
D = D';
E = [-C;ones(1,n-1)];
```

```
c = inv(E * E') * E * D;
c = c';
a = c(1);b = c(2);
% 预测后续数据
F = [];F(1) = A(1);
for i = 2:(n + 10)
    F(i) = (A(1) - b/a)/exp(a * (i - 1)) + b/a ;
end
G = [];G(1) = A(1);
for i = 2:(n + 10)
    G(i) = F(i) - F(i - 1);              % 得到预测出来的数据
end
t1 = 1999:2008;
t2 = 1999:2018;
G
plot(t1,A,'o',t2,G)                  % 原始数据与预测数据的比较
```

运行该程序,得到的预测数据如下:

```
G =
  1.0e + 006 *
  Columns 1 through 14
     0.0897     0.0893     0.1034     0.1196     0.1385     0.1602     0.1854
0.2146     0.2483     0.2873     0.3325     0.3847     0.4452     0.5152
  Columns 15 through 20
  0.5962     0.6899     0.7984     0.9239     1.0691     1.2371
```

该程序还显示了预测数据与原始数据的比较图,如图 3-1 所示。

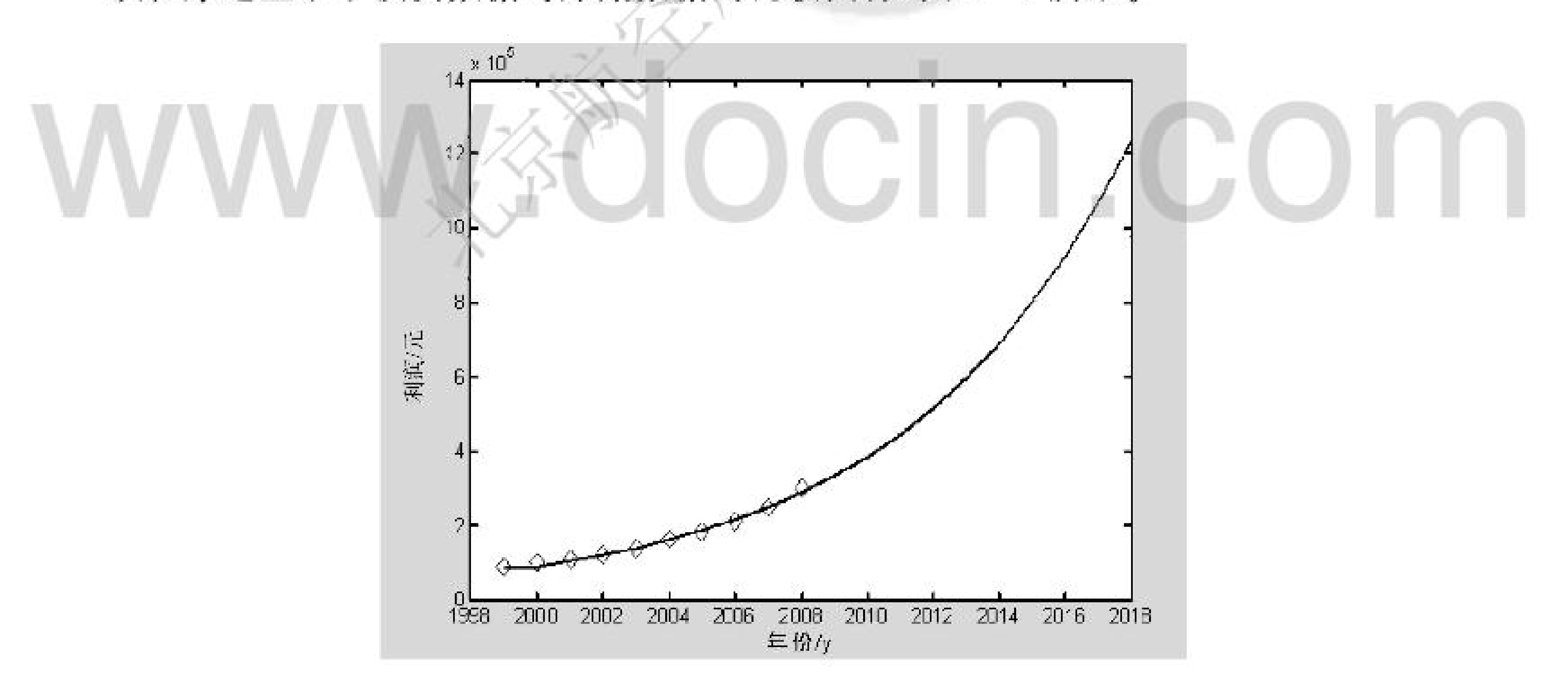


图 3-1　某公司利润预测数据与原始数据的比较

3.2.2　灰色预测程序说明

(1) 先熟悉程序中各条命令的功能,以加深对灰色预测理论的理解。

(2) 在实际使用时,可以直接套用该段程序,把原数据和时间序列数据替换就可以了。

(3) 模型的误差检验可以灵活处理,图 3-1 给出的是预测数据与原始数据的比较图,读者也可以对预测数据进行其他方式的精度检验。

3.3 灰色预测应用实例

3.3.1 实例一:长江水质的预测(CUMCM 2005 A)

长江的水质问题是一个复杂的非线性系统,但是由于数据样本少,需要预测的时间长,直接应用神经网络很难取得理想的效果。考虑到污水排放量的变化规律是一个不确定的系统,且本题给出污水排放量数据样本比较少,还要求做出长达 10 年的预测,因此采用灰色预测方法来预测未来的污水排放量。

对原题附件 4 中的数据进行整理可以得到 10 年的长江污水量排放数据,如表 3-3 所列。

表 3-3　1995—2004 年长江污水排放量

年　份	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
污水量/亿吨	174	179	183	189	207	234	220.5	256	270	285

以 P3-1 的程序段为基础,将表 3-3 的数据带入,并更新时间轴数据,即得到新程序 P3-2。程序输出的图形如图 3-2 所示。

程序编号	P3-2	文件名称	Changjiang.m	说明	预计长江污水排放量

```
clear
syms a b;
c=[a b]';
A=[174    179    183    189    207 234    220.5 256    270    285];
B=cumsum(A);                              %原始数据累加
n=length(A);
for i=1:(n-1)
    C(i)=(B(i)+B(i+1))/2;                 %生成累加矩阵
end
% 计算待定参数的值
D=A;D(1)=[];
D=D';
E=[-C;ones(1,n-1)];
c=inv(E*E')*E*D;
c=c';
a=c(1);b=c(2);
% 预测后续数据
F=[];F(1)=A(1);
for i=2:(n+10)
    F(i)=(A(1)-b/a)/exp(a*(i-1))+b/a ;
end
G=[];G(1)=A(1);
for i=2:(n+10)
    G(i)=F(i)-F(i-1);                     %得到预测出来的数据
```

```
end
t1 = 1995:2004;
t2 = 1995:2014;
G;a, b                          % 输出预测值、发展系数和灰色作用量
plot(t1,A,'o',t2,G)             % 原始数据与预测数据的比较
```

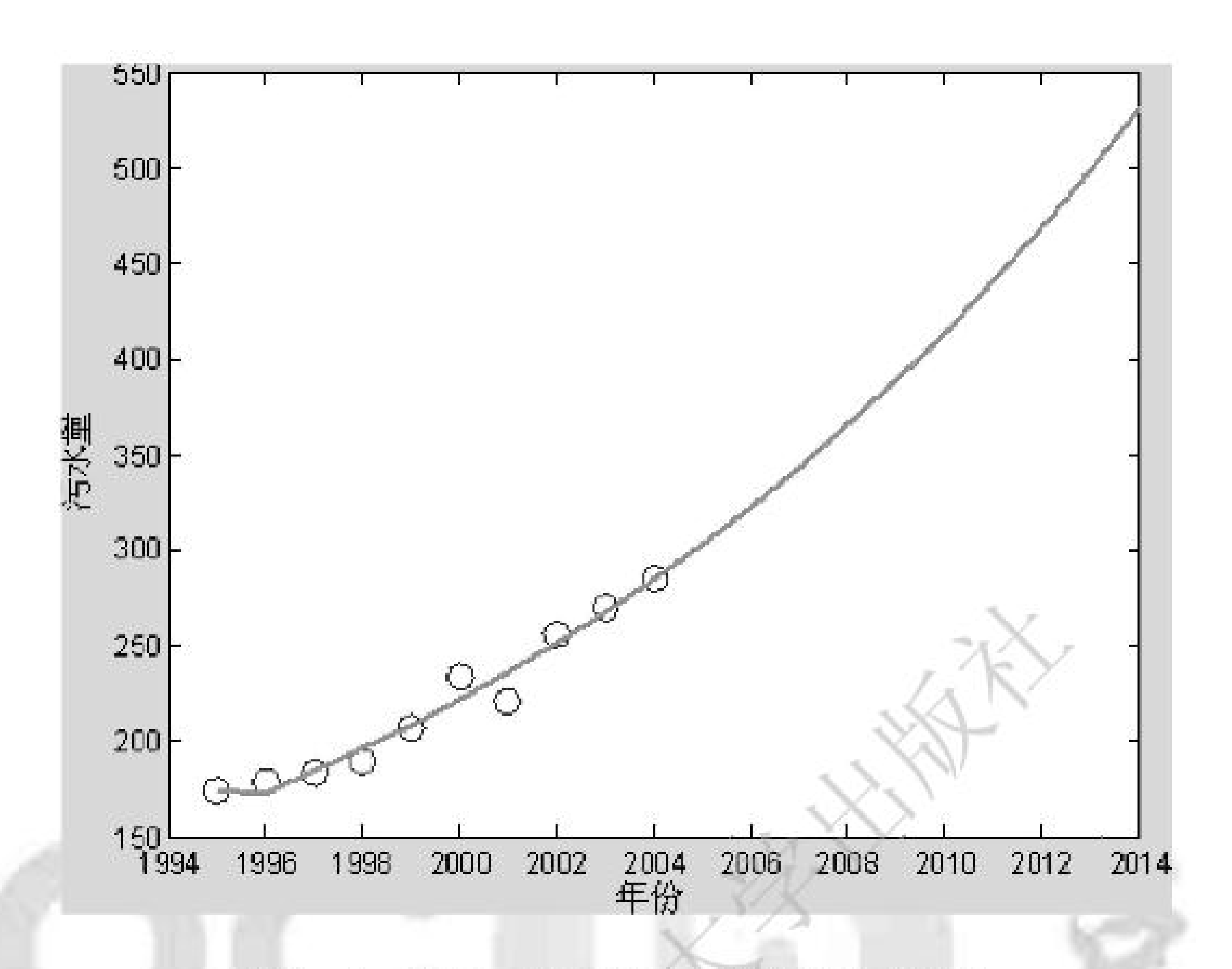


图 3-2　1995—2014 预计长江排放的污水量

3.3.2　实例二：预测与会代表人数(CUMCM 2009 D)

1. 问题描述

该题要求为会议筹备组制订一个预订宾馆客房、租借会议室、租用客车的合理方案。为了解决这个问题，需要先预测与会代表的人数。预测的依据是代表回执数量及往届的与会人员数据。已知本届会议的回执情况(见表 3-4)及以往几届会议代表回执和与会情况(见表 3-5)。要解决的问题是：根据这些数据预测本届与会代表人数。

表 3-4　本届会议代表的回执中有关住房要求的信息　　单位：人

要　求	合住 1	合住 2	合住 3	独住 1	独住 2	独住 3
男	154	104	32	107	68	41
女	78	48	17	59	28	19

说明：表头第一行中的数字 1、2、3 分别指每天每间 120～160 元、161～200 元、201～300 元三种不同价格的房间。合住是指要求两人合住一间。独住是指可安排单人间或一人单独住一个双人间。

表 3-5　以往几届会议代表回执和与会情况　　单位：人

届　次	第一届	第二届	第三届	第四届
发来回执的代表数量	315	356	408	711
发来回执但未与会的代表数量	89	115	121	213
未发回执而与会的代表数量	57	69	75	104

2. 问题的求解

根据表 3－4 的数据，可知本届发来回执的数量为 755。根据表 3－5 的数据，可以知道发来回执但未与会的代表数和未发回执但与会的代表与发来回执数量间的关系。

定义 1：未知与会率＝未发回执但与会的代表的数量/发来回执的代表数量

定义 2：缺席率＝发来回执但未与会的代表数/发来回执的代表数量

根据以上定义，可以得到往届的缺席率和未知与会率，如表 3－6 所列。

表 3－6　往届的缺席率和未知与会率

届　次	第一届	第二届	第三届	第四届
缺席率	0.282540	0.323034	0.296569	0.299578
未知与会率	0.180952	0.193820	0.183824	0.146273

从表 3－6 可以看出，缺席率一直保持在 0.3 左右，而未知与会率却变化较快。为此，认为第五届的缺席率仍为 0.3。这样，缺席的人数为 755 人×0.3＝226.5 人。为了保守起见，对 226.5 人进行向下取整，即缺席的人数为 226 人。

未知与会率变化相对剧烈，不适合应用比例方法确定，同时由于数据有限，所以应用灰色预测方法比较合适。从实际问题的角度，认为以未知与会率为研究对象较为合适。将往届的未知与会率数据带入程序 P3－1，并对输入数据和预测数据做相应修改，可很快得到本届的未知与会率为 0.1331，所以本届未发回执但与会的代表数量为 755 人×0.1331＝ 100.4905 人。同样保守考虑，向上取整为 101 人。这样就可以预测本届与会代表的数量为 755 人＋101 人－226 人＝630 人。

第 4 章

遗传算法及其 MATLAB 实现

4.1 遗传算法基本原理

4.1.1 人工智能算法概述

前面已经学习了数学建模中常规方法的使用、不太复杂问题的 MATLAB 规划工具箱的求解以及灰色预测理论。从本章开始，将陆续接触一些常见的、相对复杂的智能算法和其他一些高级方法的基本理论并用 MATLAB 软件实现。

海量数据、信息高速公路和知识爆炸都是现代社会三个重要的基本属性，数据、信息和知识是数字时代的关键词。尽管知识离不开数据和信息，但知识与数据和信息并不是同一概念。根据美国著名管理学家 Richard L. Daft 的经典著作《组织理论与设计》的释义，数据(data)是简单的、纯粹的事实和数字，其本身包含的使用价值很小。比如，一个企业可能有项数据显示出其生产的某种产品有 20%销往江苏省。要想这个数据有用，就必须将其与其他数据横向或纵向联系起来，如“这批产品被 16～25 岁的青年男女买走”，将其处理为信息。所谓信息(information)就是与其他的数据联系后所转化成的具有特定用途的具有使用价值的数据。知识(knowledge)就更进一步了，它是指将某一信息与其他信息相联系，并与已知的信息进行比较后而从这一信息中得出的结论。与信息和数据对比，知识总是包含着人或者智能的因素。书本能承载信息，但是这些必须被人吸收并付诸使用后才能成为知识。书本上的知识是建立在已有的信息、亲身体验、直觉和理解的基础上的，它包含着所掌握的信息应如何行动的认知能力。

智能算法是智能技术领域的一个分支。智能算法出现的原因是，人们生活在知识新陈代谢速度快和信息繁杂的社会里，需要用高效的数据挖掘工具从自然人的各类密码、账号、消费习惯所形成的数据中提取出有益的信息和知识以便为产品的销售提供决策支持，以及从企业的各种海量数据报表中筛选出信息和知识以便提高生产效率降低成本，或者用计算机代替人脑进行复杂的各类计算。如何从海量数据中采集到信息，又如何从信息中归纳知识，是一项高技术并且劳动强度很大的工作。为了减轻人们的劳动量，同时也为了节约社会成本，出现了一些载有能代替人类脑力劳动的算法的机器硬件，这些算法被称为智能算法，它们具有一些类似生命体智慧的特征，其实就是经过抽象以后的机器学习。不论是粒子群算法还是蚁群算法，也不论是遗传算法还是免疫算法，更不论是神经网络还是模拟退火算法，本质上都是机器学习。机器学习具有一个显著特征——机械性。吴文俊院士在《数学机械化》(科学出版社，2003 年 3 月)一书中提到，数学机械化是指数学在运算和证明过程中，只要前进了一步，都会有确定的、必然的下一步，直到到达结论，整个过程是按照既定的刻板规律进行。现在流行的用计算机将解析几何转化成向量进行证明就是 20 世纪 70—80 年代吴文俊院士首创的，这是数学机械化的一个成功应用。可见，智能算法、机器学习和数学机械化在本质上是相通的，可以被看成同

一个问题的不同侧面。

智能算法是一门边缘交叉学科，是生物、数学等多学科的完美融合。各学科交叉之处往往展现出最强劲的发展势头。恩格斯在《自然辩证法》里就说过："在分子科学和原子科学的接触点上，物理学家和化学家都承认自己无能为力，然而应当在这点上期待最大的成果。"这里常常会形成一门新的科学，是"新科学"的"生长点"，是最具生命力、最活跃的地方。例如，蜡烛火焰的温度总在略带淡蓝色的外焰温度最高，焰心处温度最低；台风总在边缘处风力最大，风眼的风力为零级；靠近锅边的水总是最先沸腾；电荷在导体的尖端边缘的地方容易聚集；海岸是陆地和海洋作用最强烈的地方；磁铁两端的 N/S 极总是磁性最强。

现代的智能算法已经很多了，不同的智能算法之间相互借鉴，不断融合，差距越来越小。

4.1.2 遗传算法生物学基础

在一定的时间里，有一群兔子，其中一些比另外一些兔子跑得快，而且更聪明，这些兔子被狐狸吃掉的可能性比较小，因此它们中的多数就存活下来并繁殖更多的兔子。当然，一些跑得慢而愚蠢的兔子也会存活下来，只是因为它们比较侥幸，这些存活的兔子群开始生育。生育的结果是兔子遗传材质的充分融合：一些跑得慢的兔子生出了跑得快的兔子，一些跑得快的兔子生出跑得更快的，一些聪明的兔子生出了愚蠢的兔子，等等。在最顶层，自然界不时地变异一些兔子的基因材质。所产生的小兔子平均来说要比原始的群体更快更聪明，因为从狐狸口中生存下来的父代多数是跑得更快、更聪明的兔子。同样，狐狸也经历相似的过程，否则兔子可能跑得太快又太聪明以致狐狸根本抓不到了。

兔子的生存哲学就是以自然选择学说为核心的现代生物进化理论，其基本观点是：种群是生物进化的基本单位，生物进化的实质是种群基因频率的改变。基因突变和基因重组、自然选择及隔离是物种形成过程的三个基本环节，通过它们的综合作用，种群产生分化，最终导致新物种的形成。在这个过程中，基因突变和基因重组产生生物进化的原材料，自然选择使种群的基因频率定向改变并决定生物进化的方向，隔离是新物种形成的必要条件。

新物种形成的途径和方式有两种：渐变式和爆发式。渐变式主要通过变异的逐渐积累而成亚种，再由亚种形成一个或多个新种，新种又分为两种类型，即继承式新种形成和分化式新种形成；爆发式不通过亚种这一阶段而迅速形成新的物种，其分为三种类型，即杂交产生新种，染色体结构变化形成新种和多倍体化的新种形式。遗传算法杂交了渐变式和爆发式的两种思想。

4.1.3 遗传算法的实现步骤

1975 年 Michigan 大学的 J. Holland 教授提出"遗传算法"的概念并出版了颇有影响的著作《Adaptation in Natural and Artificial Systems》。遗传算法是模拟达尔文生物进化论的自然选择和孟德尔遗传学机理的生物进化过程的计算模型，是一种通过模拟自然进化过程搜索最优解的方法。J. Holland 教授提出的遗传算法属于简单遗传算法(simple genetic algorithm ,SGA)，还处于遗传算法发展进程的初级阶段，但却具有里程碑式的意义。

SGA 由编解码、个体适应度评估和遗传运算三大模块构成，而遗传运算又包括染色体复制、交叉、变异甚至倒位等。改良的遗传算法和融合新型技术的遗传算法都是 SGA 的变异形式。在遗传算法中，定义种群或群体为所有编码后的染色体集合，表征每个个体的是其相应的染色体。

1. 编　码

遗传算法的编码有浮点编码和二进制编码两种，这里只介绍二进制编码规则。二进制编码既符合计算机处理信息的原理，也方便了对染色体进行遗传、编译和突变等操作。设某一参数的取值范围为(L,U)，使用长度为k的二进制编码表示该参数，则它共有2^k种不同的编码。该参数编码时的对应关系为

$$000000000000000000=0\rightarrow L$$
$$000000000000000001=1\rightarrow L+\delta$$
$$000000000000000010=2\rightarrow L+2\delta$$
$$000000000000000011=3\rightarrow L+3\delta$$
$$\cdots\cdots$$
$$111111111111111111=2^k-1\rightarrow U$$

易知：

$$\delta=\frac{U-L}{2^k-1}$$

2. 解　码

解码的目的是为了将不直观的二进制数据串还原成十进制。设某一个体的二进制编码为$b_kb_{k-1}b_{k-2}\cdots b_3b_2b_1$，则对应的解码公式为

$$x=L+\left(\sum_{i=1}^{k}b_i2^{i-1}\right)\frac{U-L}{2^k-1}$$

例如，设有参数$x\in[2,4]$，现用5位二进制数对x进行编码，可得$2^5=32$条染色体：

00000，00001，00010，00011，00100，00101，00110，00111
01000，01001，01010，01011，01100，01101，01110，01111
10000，10001，10010，10011，10100，10101，10110，10111
11000，11001，11010，11011，11100，11101，11110，11111

对于任意二进制数据串只要代入译码公式，就可以得到对应的解码，如$x_{22}=10101$，它对应的十进制值为

$$\sum_{i=1}^{5}b_i2^{i-1}=1\times2^0+0\times2+1\times2^2+0\times2^3+1\times2^4=21$$

则对应参数的x值为

$$x=2+21\times\frac{4-2}{2^5-1}=3.3548$$

遗传算法的编码和解码在宏观上可以对应生物的基因型和表现型，在微观上可以对应DNA的转录和翻译两个过程。

3. 交　配

“交配运算”是使用单点或多点进行交叉的算子。首先用随机数产生一个或多个交配点位置，然后两个个体在交配点位置互换部分基因码，形成两个子个体。例如，有两条染色体$S_1=01001011$，$S_2=10010101$交换其后4位基因，如图4-1所示。$S'_1=01000101$，$S'_2=10011011$可以被看做是原染色体S_1和S_2的子代染色体。

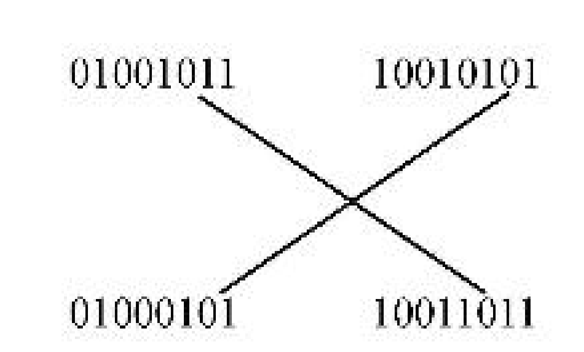


图4-1　染色体基因交配范例

4. 突　变

“突变运算”是使用基本位进行基因突变。为了避免在算法迭代后期出现种群过早收敛，对于二进制的基因码组成的个体种群，实行基因码的小几率翻转，对于二进制编码即 0 变为 1，而 1 变为 0。例如，将染色体 $S=11001101$ 第 3 位上的 0 变为 1，即 $S=11\bar{0}01101 \rightarrow 11\bar{1}01101=S'$。$S'$可以被看做是原染色体 S 的子代染色体。

5. 倒　位

除了交配和突变之外，对于复杂的问题可能需要用到“倒位”，其对应的运算亦被称为“倒位运算”。倒位是指一个染色体某区段正常排列顺序发生 180°的颠倒，造成染色体内的 DNA 序列重新排列，它包括臂内倒位和臂间倒位。倒位纯合体不影响个体的生活力，只是改变了染色体上的相邻基因位置，从而某些表现型发生位置效应，同时也改变了与相邻基因的交换值。倒位杂合体则不然，其生育力降低。染色体上的区段可能一次又一次发生倒位，且通过自交出现不同的倒位纯合体，致使它们与其原来的物种不能交配，形成生殖隔离，结果产生新族群或变种。

倒位运算是与倒位概念相似的运算规则。例如，染色体 S'就是 S 经过倒位运算以后得到的染色体基因编码，其中倒位的部分即 S 中标记下画线的部分。

$$S = 10010110\underline{11101110011010}101001 \Rightarrow S' = 10010110\underline{01011001110111}101001$$

6. 个体适应度评估

自然界中能够适应环境的生物有更多的机会存活下来，这种筛选机制类似图 4－2 所示现象：在由正六边形搭建的三角形区域的顶部投掷一些光滑的木块，这些木块经由白色的缝隙坠落底部，显然落在底部中间的木块要比落在两端的木块多，因为木块有更多的路径坠落在底部的中间区域，所以有更大的几率落在中间。落在各个区域的几率对应遗传算法中各条染色体被遗传到下一代的几率，其坠落的位置对应自变量取值。遗传算法依照与个体适应度成正比的几率决定当前种群中各个个体遗传到下一代群体中的机会。个体适应度大的个体更容易被遗传到下一代。通常，求目标函数最大值的问题可以直接把目标函数作为检测个体适应度大小的函数。

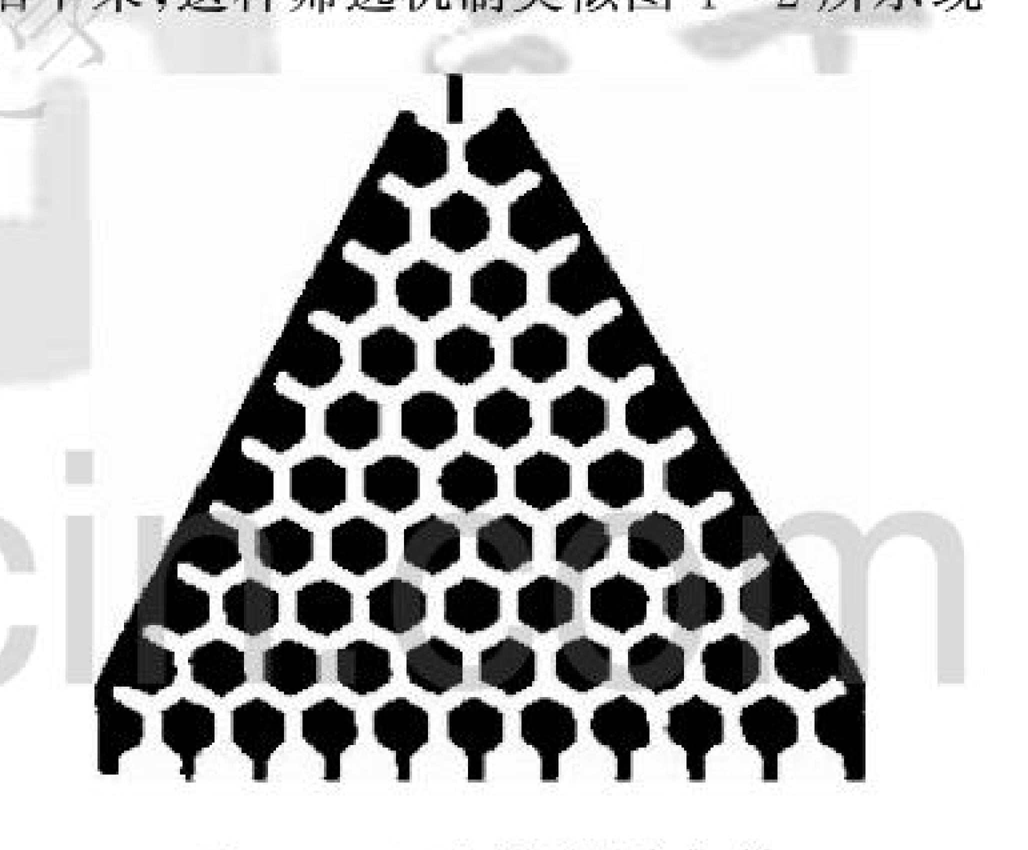


图 4－2　正六边形筛选机制

7. 复　制

复制运算是根据个体适应度大小决定其下代遗传的可能性。若设种群中个体总数为 N，个体 i 的适应度为 f_i，则个体 i 被选取的几率

$$P_i = \frac{f_i}{\sum_{k=1}^{N} f_k}$$

当个体复制的几率决定后，再产生[0,1]区间的均匀随机数来决定哪些个体参加交配。若个体适应度高，则被选取的几率 P_i 就大，则可能被多次选中，它的遗传基因就会在种群中扩散；若个体的复制几率小，则会被逐渐淘汰。

下面以一个经典案例具体说明遗传算法的作业过程。

【例 4-1】 求 $\max f(x_1,x_2)=21.5+x_1\sin(4\pi x_1)+x_2\sin(20\pi x_2)$

$$\text{s.t.}\begin{cases}-3.0\leqslant x_1\leqslant 12.1\\ 4.1\leqslant x_2\leqslant 5.8\end{cases}$$

函数 $f(x_1,x_2)$ 的三维图形如图 4-3 所示。

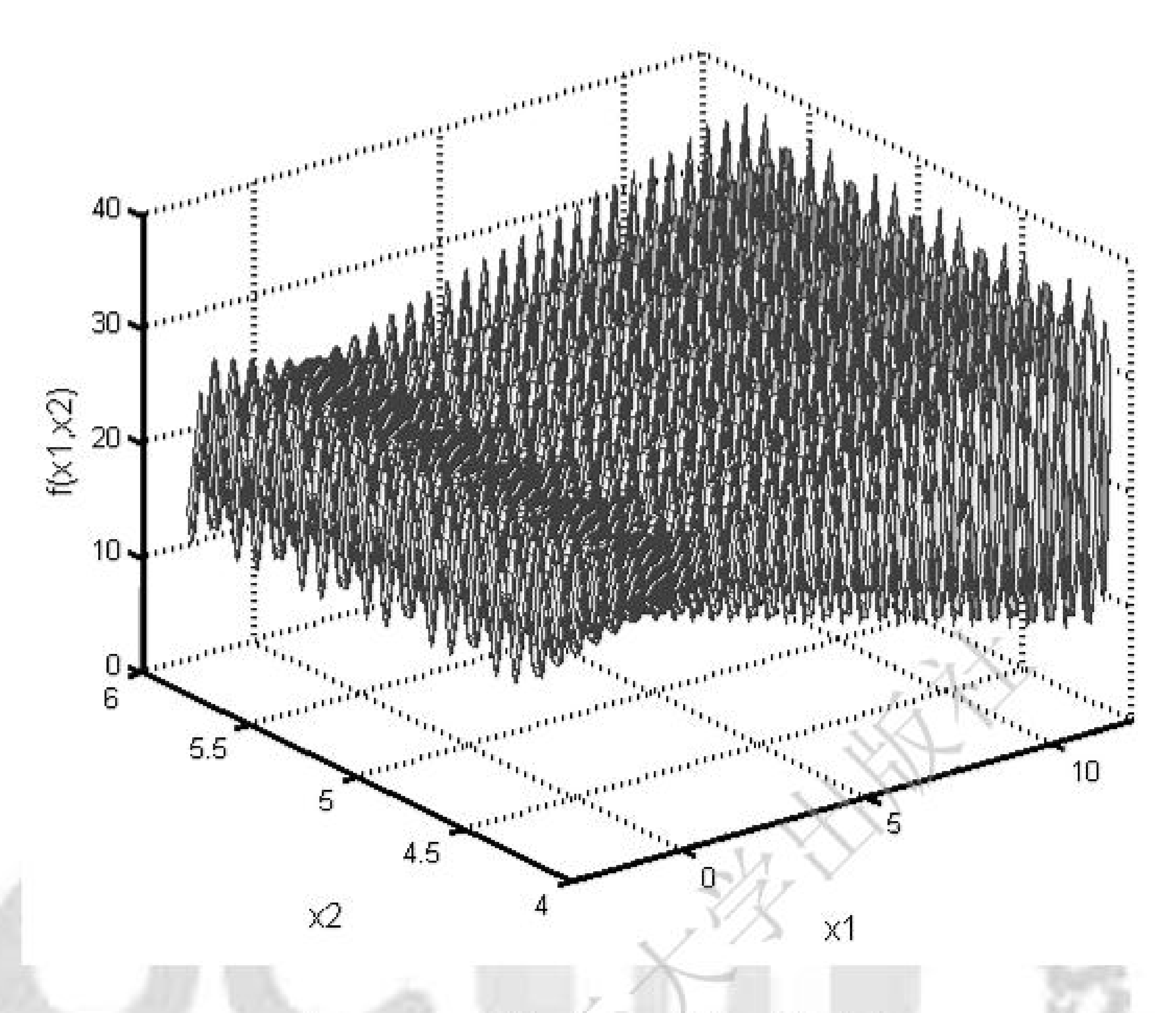


图 4-3　函数 $f(x_1,x_2)$ 的三维图形

(1) 编　码

首先要进行编码工作，即将变量转换成二进制数串。数串的长度取决于所要求的精度。例如，变量 x 的区间是 (L,U)，要求的精度是小数点后 4 位，也就意味着每个变量应该被分成至少 $(L,U)\times 10^4$ 个部分。对一个变量的二进制数串位数用以下公式计算：

$$2^{m_j-1}<(U-L)\times 10^4\leqslant 2^{m_j}-1$$

本例精度要求保留小数点后 4 位，则目标函数的两个自变量 x_1 及 x_2 所构成的染色体数串可以表示如下：

$$\begin{cases}-3.0\leqslant x_1\leqslant 12.1\\ [12.1-(-3.0)]\times 10^4=151\,000\\ 2^{m_1-1}<151\,000\leqslant 2^{m_1}-1\\ m_1=18\\ 4.1\leqslant x_2\leqslant 5.8\\ (5.8-4.1)\times 10^4=17\,000\\ 2^{m_2-1}<17\,000\leqslant 2^{m_2}-1\\ m_2=15\\ m=m_1+m_2=18+15=33\end{cases}$$

本例中任一染色体数串都是 33 位，即 000001010100101001101111011111110。

以上编码前 18 位表示 x_1，后 15 位表示 x_2，如表 4-1 所列。

表 4-1　染色体编码

自变量	二进制	十进制	实际值
x_1	000001010100101001	5417	−2.687969
x_2	101111011111110	24318	5.361653

则二进制转化成十进制为

$$x_1 = -3.0 + 5417 \times \frac{12.1-(-3.0)}{2^{18}-1} = -2.687969$$

$$x_2 = 4.1 + 24318 \times \frac{5.8-4.1}{2^{15}-1} = 5.361653$$

假设初始种群中有 10 个个体，其染色体可随机生成如下：

$U_1 = [000001010100101001\ 101111011111110]$
$U_2 = [001110101110011000\ 000010101001000]$
$U_3 = [111000111000001000\ 010101001000110]$
$U_4 = [100110110100101101\ 000000010111001]$
$U_5 = [000010111101100010\ 001110001101000]$
$U_6 = [111110101011011000\ 000010110011001]$
$U_7 = [110100010111110001\ 001100111011101]$
$U_8 = [001011010100001100\ 010110011001100]$
$U_9 = [111110001011101100\ 011101000111101]$
$U_{10} = [111101001110101010\ 000010101101010]$

相对应的十进制的实际值$[x_1, x_2]$为

$U_1 = [-2.687969, 5.361653]$，　$U_2 = [0.474101, 4.170144]$
$U_3 = [10.419457, 4.661461]$，　$U_4 = [6.159951, 4.109598]$
$U_5 = [-2.301286, 4.477282]$，　$U_6 = [11.788084, 4.174346]$
$U_7 = [9.342067, 5.121702]$，　$U_8 = [-0.330256, 4.694977]$
$U_9 = [11.671267, 4.873501]$，　$U_{10} = [11.446273, 4.171908]$

（2）评价个体适应度

对一个染色体数串的适应度的评价由下列三个步骤组成。

① 将染色体串进行反编码（解码），转换成真实值，即

$$x^k = (x_1^k, x_2^k),\qquad k = 1,2,3,\cdots$$

② 评价目标函数 $f(x^k)$。

③ 将目标函数值转为适应度。对于极大值问题，适应度可作为目标函数值：

$$\text{eval}(U_k) = f(x^k),\qquad k = 1,2,3,\cdots$$

在遗传算法中，评价函数扮演自然进化中环境的角色，它通过染色体的适应度对其进行评价。上述染色体的适应度值如下：

$$\text{eval}(U_1) = f(-2.687969, 5.361653) = 19.805119$$
$$\text{eval}(U_2) = f(0.474101, 4.170144) = 17.370896$$
$$\text{eval}(U_3) = f(10.419457, 4.661461) = 9.590546$$
$$\text{eval}(U_4) = f(6.159951, 4.109598) = 29.406122$$

$$\text{eval}(U_5) = f(-2.301286, 4.477282) = 15.686091$$
$$\text{eval}(U_6) = f(11.788084, 4.174346) = 11.900541$$
$$\text{eval}(U_7) = f(9.342067, 5.121702) = 17.958717$$
$$\text{eval}(U_8) = f(-0.330256, 4.694977) = 19.763190$$
$$\text{eval}(U_9) = f(11.671267, 4.873501) = 26.401669$$
$$\text{eval}(U_{10}) = f(11.446273, 4.171908) = 10.252480$$

依照染色体的适应度值进行新种群的复制，步骤如下：

① 计算染色体 U_k 的适应度值

$$\text{eval}(U_k) = f(x^k), \qquad k = 1, 2, \cdots$$

② 计算种群的适应度值总和

$$F = \sum_{k=1}^{\text{pop_size}} \text{eval}(U_k)$$

③ 计算每个染色体被复制的概率

$$P_k = \frac{\text{eval}(U_k)}{F}$$

④ 计算每个染色体被复制的累积概率

$$Q_k = \sum_{j=1}^{k} P_k$$

(3) 新种群复制

依照轮盘选择法，转动轮盘 10 次(种群中有 10 条染色体)，每次选择一个作为新种群的染色体。假设 10 次中产生的 0～1 随机数序列如下：

0.301431　0.322062　0.766503　0.881893　0.350871

0.538392　0.177618　0.343242　0.032685　0.197577

根据以上的计算方法，可以先计算出种群中每个染色体的适应度和概率，如表 4-2 所列。利用计算机模拟轮盘选择法，假设计算机产生 10 个[0,1]区间的随机数列如下：

表 4-2　种群每条染色体的适应度、被复制概率和被复制的累积概率

染色体	适应度值	P_k	Q_k
U_1	19.805119	0.111180	0.111180
U_2	17.370896	0.097515	0.208695
U_3	9.590546	0.053839	0.262534
U_4	29.406122	0.165077	0.427611
U_5	15.686091	0.088057	0.515668
U_6	11.900541	0.066806	0.582475
U_7	17.958717	0.100815	0.683290
U_8	19.763190	0.110945	0.794234
U_9	26.401669	0.148211	0.942446
U_{10}	10.252480	0.057554	1.000000

第 1 个随机数为 0.301431，大于 Q_3 小于 Q_4，所以 U_4 被选中；

第 2 个随机数为 0.322062，大于 Q_3 小于 Q_4，所以 U_4 再次被选中；

第 3 个随机数为 0.766 503，大于 Q_7 小于 Q_8，所以 U_8 被选中；

……

第 10 个随机数为 0.197 577，大于 Q_1 小于 Q_2，所以 U_2 被选中；

依照轮盘选择法，新种群的染色体组成如下：

$$U_1 = [100110110100101101\ 000000010111001](U_4)$$
$$U_2 = [100110110100101101\ 000000010111001](U_4)$$
$$U_3 = [001011010100001100\ 010110011001100](U_8)$$
$$U_4 = [111110001011101100\ 011101000111101](U_9)$$
$$U_5 = [100110110100101101\ 000000010111001](U_4)$$
$$U_6 = [110100010111110001\ 001100111011101](U_7)$$
$$U_7 = [001110101110011000\ 000010101001000](U_2)$$
$$U_8 = [100110110100101101\ 000000010111001](U_4)$$
$$U_9 = [000001010100101001\ 101111011111110](U_1)$$
$$U_{10} = [001110101110011000\ 000010101001000](U_2)$$

这种轮盘选择法的机理是：染色体的适应度大意味着$[Q_k, Q_{k+1}]$区间跨度就大，随机数发生器产生的均匀随机数就会有更大的概率落在较大长度的$[Q_k, Q_{k+1}]$区间里，这样具有较大 P_k 值的染色体自然更有机会复制到下一代。

(4) 新种群交配

① 交配染色体数量的确定

交配染色体的数量等于染色体总量乘以交配概率。这里假设交配概率 P_c 为 0.25，染色体总量为 10 条，所以参加交配的染色体数量为[2.5]条。符号[ ]表示取整，这里取整数 2，即交配的染色体数目为 2 条。

② 交配染色体对象的确立

用计算机产生[0,1]区间的 10 个随机数如下：

0.625 721	0.266 823	0.288 644	0.295 114	0.163 274
0.567 461	0.085 940	0.392 865	0.770 714	0.548 656

假定其分别对应 $U_1 \sim U_{10}$ 这 10 个个体，则其中低于交配概率 0.25 的 U_5 和 U_7 参加交配。这样操作的原因是：交配概率越低，低于交配概率以下的随机数的数量就越少，所以参加交配的染色体数量与交配概率可能会成正比。

③ 在交配池发生交配

染色体 U_5 和 U_7 被选中作为交配的父辈，交配点的选择以随机数产生。交配的种类有单点交配和多点交配，这里取单点交配。计算机随机生成一个介于 0～32 的整数(因为整个染色体数串的长度为 33)。假设所产生的整数为 1，那么两个染色体自 1 位置(即二进制串的第二位)开始分割，在染色体 1 位置右端部分进行交换而生成新的子辈染色体，即

$$U_5 = [1\ \underline{0011\ 0110\ 1001\ 0110\ 1000\ 0000\ 1011\ 1001}]$$
$$U_7 = [0\ \underline{\underline{0111\ 0101\ 1100\ 1100\ 0000\ 0101\ 0100\ 1000}}]$$
$$\Downarrow$$
$$U_5^* = [1\ \underline{\underline{0111\ 0101\ 1100\ 1100\ 0000\ 0101\ 0100\ 1000}}]$$
$$U_7^* = [0\ \underline{0011\ 0110\ 1001\ 0110\ 1000\ 0000\ 1011\ 1001}]$$

(5) 基因突变

依照突变运算规则并假设突变几率 P_m 为 0.01，亦即种群内所有基因都有 0.01 的概率进行突变。在本例中共有 33×10=330 个基因，即希望每一代中有 3.3 个突变基因，每个基因的突变几率是均等的。因此，将产生 330 个介于 0～1 之间的随机数(需编号)，然后将该随机数小于 0.01 者选出，并将其对应的基因值加以翻转，假设 330 次中产生 0～1 之间的随机数，其值小于 0.01 者如表 4-3 所列。

表 4-3　基因突变位置

基因位置	染色体位置	基因位数	随机数
105÷33=4…6	4	6	0.009857
164÷33=5…32	5	32	0.003113
199÷33=7…1	7	1	0.000946
329÷33=10…32	10	32	0.001282

表 4-3 第 1 列显示的是具体在哪些染色体以及在染色体的什么位置进行了突变。例如第 1 行"105÷33=4…6"表示的是在第 4 条染色体第 6 个基因上发生了突变，因为第 4 条染色体第 6 个基因对应的基因编号是 33×(4−1)+6=105。

在突变后，最终新种群的染色体组成如下：

$$U_1 = [100110110100101101\ 000000010111001]$$
$$U_2 = [100110110100101101\ 000000010111001]$$
$$U_3 = [001011010100001100\ 010110011001100]$$
$$U_4 = [111111001011101100\ 011101000111101]$$
$$U_5 = [101110101110011000\ 000010101001010]$$
$$U_6 = [110100010111110001\ 001100111011101]$$
$$U_7 = [100110110100101101\ 000000010111001]$$
$$U_8 = [100110110100101101\ 000000010111001]$$
$$U_9 = [000001010100101001\ 101111011111110]$$
$$U_{10} = [001110101110011000\ 000010101001010]$$

新一代的相对应实际值[x_1, x_2]和适应度值如下：

$$\text{eval}(U_1) = f(6.159951, 4.109598) = 29.406122$$
$$\text{eval}(U_2) = f(6.159951, 4.109598) = 29.406122$$
$$\text{eval}(U_3) = f(-0.330256, 4.694977) = 19.763190$$
$$\text{eval}(U_4) = f(11.907206, 4.873501) = 5.702781$$
$$\text{eval}(U_5) = f(8.024130, 4.170248) = 19.910251$$
$$\text{eval}(U_6) = f(9.342067, 5.117020) = 17.958717$$
$$\text{eval}(U_7) = f(6.159951, 4.109598) = 29.406122$$
$$\text{eval}(U_8) = f(6.159951, 4.109598) = 29.406122$$
$$\text{eval}(U_9) = f(-2.687969, 5.361653) = 19.805119$$
$$\text{eval}(U_{10}) = f(0.474101, 4.170248) = 17.370896$$

至此，已完成遗传算法的第一代流程。依此迭代，在第 451 代得到对应最大目标函数值的

染色体：

$$U_{best} = [111110000000111000\ 111101001010110]$$

相应实际值$[x_1, x_2]=[11.631407, 5.724824]$，适应度值 $eval(U)=f(11.631407, 5.724824)=38.818208$

从以上实例可以得出两点结论：

① 遗传算法本质上是一种启发式的随机搜索算法，所以由遗传算法得出的结果每次都不尽相同。

② 自变量在给定的约束条件下进行了无缝编码(即这种编码方式能够表达解空间中的所有可行解)，所以从理论上讲，遗传算法总有很多机会得到全局最优结果而不是局部最优结果。

4.1.4 遗传算法的拓展

1. 协同进化遗传算法

遗传算法是建立在生物进化理论和染色体遗传变异基础之上的，该理论遵循优胜劣汰的自然选择机制，本质是“胜者为王，败者为寇”的生物种群内部激烈的竞争关系。但实际上，生物在进化过程中，除了竞争关系，还有协作、寄生等关系。生物与生物之间、生物与环境之间存在着“相似相溶和相反相成”的和谐共处关系。所以，基于达尔文进化论和孟德尔遗传定律的遗传算法不可避免地存在着弱点，这为标准遗传算法处理异常复杂的优化问题时显得无能为力得到佐证。

协同进化的概念最早由 Ehrlich 和 Raven 讨论植物和植食昆虫(蝴蝶)相互之间的进化影响时提出的。借鉴生态系统中的协同和协作关系，Jason 提出了协同进化计算机制，并给出了系统进化算法的具体数学模型。相关协同进化理论和 MATLAB 源程序可参阅文献[4]。

2. 遗传算法与弗洛伊德梦的解析法

一般认为梦是这样形成的：白天，人的左大脑半球工作最忙，它从感觉器官中接受信息，进行加工，解决可能出现的各种问题。同时，右大脑半球也悄悄地工作，神经元记录下那些自己来不及考虑的情绪和信息。当人们熟睡后，切换到右大脑“主持”工作，白天印在右半球里的全部潜意识以虚拟梦的形式进行释放。

如果将梦的成形过程进行离散，梦由以下程序组建而成：梦的改装、梦的置换、梦的替代、梦的凝缩、梦的转移、梦的表现、梦的仿同、梦的颠倒和梦的选择，经由这些步骤对梦思进行编译，编译后的梦内容一般难以识别，而后通过大脑的整合作用将这些代码碎片进行链接，形成连贯的梦的情节，这就是基于弗洛伊德思维的梦的建模过程。译梦则是梦的成形过程的逆运算。

着重简介一下“梦的仿同”成形原理。梦的仿同是指将梦思中的具有某一相同属性的人或物的性状都集于某一人或某一物上，这个过程称为仿同(集锦)。仿同是用在人身上，而集锦则是对事物的统一。不过集锦亦可施用在人身上。在仿同作用里，只有和共同元素相连的人才能表现在梦的显意里，其他人则被抑制了。这跟孟德尔的遗传定律原理是一样的，当显性基因存在的时候，隐性基因就不能表现出隐性性状，除非在完全隐性的环境里。梦在形成过程中都要经过审查制度。一些有违共识和有损伦理常纲的人和物的属性是逃不过梦的审查的。尽管人在潜意识里都有兽性的一面，在利比多的驱使下都有原始冲动的倾向，但是人类在后天的学习过程中受到了文化、舆论、礼仪、法律、传统等方面的教化和约束，在实际生活中形成了牢固的自我约束意识，即便在梦中这种意识也没有消除，表现为梦的审查制度。在梦中如果不能逃

脱梦的审查制度的素材演变成隐性，与另一个能够逃脱审查制度的人或物集成在一起，如此便能逃脱审查，这就是仿同(集锦)在梦形成中所起的作用。梦在形成过程中，就这样伴随着审查与欺骗的相互博弈。

弗洛伊德梦的解析法虽然隶属于心理学，遗传算法是人工智能技术的一个分支，表面看起来两者之间没有任何耦合关联，但是仔细探究可知二者有众多相似之处。例如，梦的置换和替代可以类比于遗传算法中染色体上基因片段的交叉和变异；梦的表现可以类比遗传算法中生物表现型向基因型映射的过程，也就是遗传算法中种群编码的过程；梦的颠倒可以类比遗传算法中倒位运算……其实梦不仅与遗传算法关系密切，如果是基于弗洛伊德的解梦理论，梦还与小波分析、神经网络、人工免疫算法等均有联系。在此将弗洛伊德梦的解析法与遗传算法联系起来，并非牵强附会，而是希望能起到抛砖引玉、举一反三的作用，希望能在梦与人工智能技术之间找到契合点，开发出一种“算法仿同”的高性能算法，希望这种算法有朝一日能“愿望达成”。事实上，不同智能算法之间已经相互仿同、相互借鉴、相互融合，有异曲同工的趋势，差距越来越小了。

4.2 遗传算法的MATLAB程序设计

4.2.1 程序设计流程及参数选取

1. 遗传算法的程序设计伪代码

```
BEGIN
     t = 0 ;                               % 遗传代数
     初始化 P(t) ;                          % 初始化种群或染色体
     计算 P(t)的适应值;
     while (不满足停止准则) do
          begin
          t = t+1 ;
          从 P(t-1)中选择 P(t) ;            % 选择
          重组 P(t) ;                       % 交叉和变异
          计算 P(t)的适应值;
          end
     END
```

2. 遗传算法的参数设计原则

在单纯的遗传算法中，也并不总是收敛，即使在单峰或单调也是如此。这是因为种群的进化能力已经基本丧失，种群早熟。为了避免种群的早熟，参数的设计一般遵从以下原则：

(1) 种群的规模

当群体规模太小时，很明显会出现近亲交配，产生病态基因，而且造成有效等位基因先天缺乏，即使采用较大概率的变异算子，生成具有竞争力高阶模式的可能性仍很小，况且大概率变异算子对已有模式的破坏作用极大。同时遗传算子存在随机误差(模式采样误差)，妨碍小群体中有效模式的正确传播，使得种群进化不能按照模式定理产生所预测的期望数量；种群规模太大，结果难以收敛且浪费资源，稳健性下降。种群规模的一个建议值为 0～100。

（2）变异概率

当变异概率太小时，种群的多样性下降太快，容易导致有效基因的迅速丢失且不容易修补；当变异概率太大时，尽管种群的多样性可以得到保证，但是高阶模式被破坏的概率也随之增大。变异概率一般取0.0001～0.2。

（3）交配概率

交配是生成新种群最重要的手段。与变异概率类似，交配概率太大容易破坏已有的有利模式，随机性增大，容易错失最优个体；交配概率太小不能有效更新种群。交配概率一般取0.4～0.99。

（4）进化代数

进化代数太小，算法不容易收敛，种群还没有成熟；代数太大，算法已经熟练或者种群过于早熟不可能再收敛，继续进化没有意义，只会增加时间开支和资源浪费。进化代数一般取 100～500。

（5）种群初始化

初始种群的生成是随机的。在初始种群的赋予之前，尽量进行一个大概的区间估计，以免初始种群分布在远离全局最优解的编码空间，导致遗传算法的搜索范围受到限制，同时也为算法减轻负担。

3. 适应度函数的调整

（1）在遗传算法运行的初期阶段

群体中可能会有少数几个个体的适应度相对其他个体来说非常高。若按照常用的比例选择算子确定个体的遗传数量时，则这几个相对较好的个体将在下一代群体中占有很高的比例，在极端情况下或当群体规模较小时，新的群体甚至完全由这样的少数几个个体组成。这时交配运算就起不了什么作用，因为相同的两个个体不论在何处发生交叉行为都永远不会产生新的个体。这样就会使群体的多样性降低，容易导致遗传算法发生早熟现象（或称早期收敛），使遗传算法所求的解停留在某一局部最优点上。因此，希望在遗传算法运行的初期阶段，算法能够对一些适应度较高的个体进行控制，降低其适应度与其他个体适应度之间的差异程度，从而限制其复制数量，以维护群体的多样性。

（2）在遗传算法运行的后期阶段

群体中所有个体的平均适应度可能会接近于群体中最佳个体的适应度。也就是说，大部分个体的适应度和最佳个体的适应度差异不大，它们之间无竞争力，都会有以相接近的概率被遗传到下一代的可能性，从而使得进化过程无竞争性可言，只是一种随机的选择过程。这将导致无法对某些重点区域进行重点搜索，从而影响遗传算法的运行效率。因此，希望在遗传算法运行的后期阶段，算法能够对个体的适应度进行适当的放大，扩大最佳个体适应度与其他个体适应度之间的差异程度，以提高个体之间的竞争性。

4.2.2 MATLAB 遗传算法工具箱

1. GA 工具箱版本

为了省略艰深难懂的遗传算法数学理论和降低程序开发的难度，MATLAB 软件已经将遗传算法命令进行了封装，做成专门的遗传算法工具箱——GA Toolbox，方便用户调用。关于遗传算法工具箱，需要说明三点：